UNIVERSITY OF DELHI

RATAN TATA
LIBRARY

DELHI SCHOOL OF ECONOMICS

✓

THE RATAN TATA LIBRARY

Cl. No B28ϕ1, N47 H7.3

Ac. No. 3545, 7732.

Date of release for loan

This book should be returned on or before the date last stamped below. An overdue charge of one anna will be levied for each day the book is kept beyond that date.

25th SESSION OF THE

INTERNATIONAL STATISTICAL INSTITUTE

SEPTEMBER 6-18, 1947
WASHINGTON, D. C.

PROCEEDINGS OF THE
INTERNATIONAL STATISTICAL CONFERENCES

VOLUME III

PART A

G. Goudswaard, Volume Editor

PRINTED AND MADE IN INDIA
IN THE EKA PRESS, CALCUTTA 35

TABLE OF CONTENTS*

TABLE DES MATIERES**

PART A

PAGE

INTRODUCTION 1

SPECIAL MEETINGS

Séances spéciales

√World Census of Agriculture and Population:
Recensement mondial de la population et de l'agriculture:

World Census of Agriculture—*Conrad Taeuber* 3
Recensement mondial de l'agriculture en 1950

The 1950 Census of the Americas—*Calvert L. Dedrick* 17
Le recensement des pays de l'Amérique en 1950

√ Sampling Theory:
La théorie de l'échantillonnage:

The Influence of Agricultural Research Statistics on the Development of Sampling Theory—*Frank Yates* 27
De l'influence des statistiques de la recherche agronomique sur le développement de la théorie de l'échantillonnage

Recent Developments in Sampling Theory in the United States —*W. G. Cochran* 40
Développements récents de la théorie de sondage aux Etats-Unis

Sequential Analysis—*Abraham Wald* 67
La méthode de score

Sampling Practice:
La pratique de l'échantillonnage:

History of the Uses of Modern Sampling Procedures—*Frederick F. Stephan* 81
Aperçu historique de l'application des méthodes modernes de sondage

Sampling of Human Populations—*Morris H. Hansen* 113
Sondages de populations humaines

Commercial Uses of Sampling—*J. Stevens Stock and Joseph R. Hochstim* 129
L'usage commercial de la méthode de sondage

Developments in Substantive Economic and Social Statistics in the United States:
Développements des statistiques économiques et sociales aux Etats-Unis:

Recent Developments in Marketing Statistics—*Ross M. Cunningham* 161
Développements récents dans l'analyse des marchés

*Papers are printed in the language in which they were presented. A résumé in French follows each paper presented in English, a résumé in English follows each paper presented in French.

**Les études sont publiées dans la langue où elles ont été présentées. Chaque étude en français est suivie d'un résumé anglais, et inversement.

Page

Program of the Census Bureau—*Philip M. Hauser* 171
Le programme du Bureau de Recensement des Etats-Unis

The Program of the United States Bureau of Labor Statistics —*Ewan Clague* 178
Le programme du Bureau des Statistiques du Travail des Etats-Unis

The Agricultural Statistics Program—*Charles F. Sarle* 193
Le programme de statistiques agricoles aux Etats-Unis

The Social Security Statistics Program—*I. S. Falk* 212
Le programme de statistiques de sécurité sociale aux Etats-Unis

STATISTICAL METHODOLOGY SECTION

Section de statistique méthodologique

Contributed Papers:
Communications:

Analyse des liaisons de probabilité—*G. Darmois* 231
Analysis of Probability Relationships

Une nouvelle méthode d'analyse factorielle—*Pierre J. Delaporte* 241
A new Method of Factor Analysis

The Means of Samples—*Corrado Gini* 258
Les moyennes des échantillons

On Random Variation in Statistical Data—*Wilhelm Winkler* 272
Les variations aléatories des données statistiques

Uses of Probability Theory in Economic and Social Sciences:
Emploi du Calcul des probabilités dans l'économie et les sciences sociales:

Possibilities and Limitations of the Application of Mathematics, and Probability Theory in Particular, in the Field of Social and Economic Phenomena (summary)—*Maurice Fréchet* 284
Les possibilités et les limites de l'application des sciences mathématiques, et en particulier du calcul des probabilités, à l'étude des phénomènes économiques et sociaux (résumé)

Statistical Inference from Non-Experimental Observations: An Economic Example—*Jacob Marschak* 289
Estimation statistique à l'aide de données non-expérimentales: Un exemple économique.

A Short-Cut Method for Distinguishing Between Rigid and Disturbed Periodicity—*Herman Wold* 302
Une méthode simple pour distinguer entre des périodicités rigides et perturbées

Contributed Papers Bearing on the Development in the United States of Statistical Methodology in Engineering Research and Industrial Quality Control:
Communications sur le développement, aux Etats-Unis, de la méthodologie statistique dans la recherche technique et dans le contrôle statistique de produits manufacturés:

The Role of a Statistical Consultant in a Research Organization —*Churchill Eisenhart* 308
Le rôle d'un consultant en statistique dans une organisation de recherche

PAGE

Experimental Design Principles in Industrial Research
—*George W. Brown* 314
Les plans d'expériences dans les travaux de recherche industrielle

Statistical Quality Control in American Industry
—*William R. Pabst, Jr.* 319
Contrôle statistique des fabrications en série dans l'industrie américaine

Standard Sampling-Inspection Procedures—*W. Allen Wallis* 331
Procédures standards d'inspection par échantillonnage

The Application of Statistical Theory to the Preparation of Industrial Specifications—*John H. Curtiss* 351
L'application de la théorie statistique à la préparation de spécifications industrielles

The Dependence of Earnings on Choice of Specifications
—*Merrill M. Flood* 358
Rapport entre les bénéfices d'une entreprise et les caractéristiques spécifiées pour les produits de manufacture

Statistical Inference:
Inférence statistique:

L'estimation statistique des paramètres—*Maurice Fréchet* 363
Statistical Estimation of Parameters

Inductive Reasoning or Inductive Behavior—Modern Conceptions Relative to Statistical Inference—*Jerzy Neyman* 423
Raisonnement inductif ou comportement inductif—Les conceptions modernes de la statistique mathématique

Experimental Design in the United Kingdom and the United States:
La théorie des plans d'expériences dans le Royaume-Uni et aux Etats-Unis:

Development of the Theory of Experimental Design—*R. A. Fisher* 434
Développement de la théorie des plans d'expériences

Application of the Theory of Experimental Design in Biology
—*George W. Snedecor* 440
Application de la théorie des plans d'expériences en biologie

Limitations and Potentialities of the Application of the Theory of Experimental Design in Sociology—*Margaret Jarman Hagood* 453
Les limitations et les possibilités de l'application de la théorie des plans d'expériences en sociologie

SOCIAL STATISTICS SECTION

Section de statistique sociale

Report & Contributed Papers:
Rapport et communications:

L'activité de la Commission Mixte de Statistique Criminelle
—*Alessandro A. Molinari* 458
Activities of the Joint Commission on Criminal Statistics

PAGE

Methods of the Statistics of Social Insurance—*Demetrious Calitsounakis*...... 468
Méthodes de statistique d'assurances sociales

Origin, Functions, and Future of the British Central Statistical Office—*Ernest H. Godfrey*.......... 471
Origine, fonction et avenir de l'Office Central Britannique de Statistique

Une famille des classes moyennes d'une guerre mondiale à l'autre (1914-1947)—*Ugo Giusti*.......... 480
A Middle Class Family from One World War to the Other (1914-1947)

Osservazioni sul Metodo di Rilevazione sul Volume della Circolazione Sociale —*Livio Livi*.......... 489
Remarques sur la méthode de déterminer la capillarité sociale
Some Remarks on the Method of Determining Social Mobility

Cost of living, Salaries, Consumption, and the Situation of the Italian Middle Classes—*Guglielmo Tagliacarne*.......... 504
Cout de la vie, salaires, consommation et condition des classes moyennes en Italie

Public Opinion Research:
Etudes de l'opinion publique:

Les progrès récents de l'étude des opinions en Europe— *Jean Stoetzel*...... 518
Recent Progress in Public Opinion Research in Europe
Public Opinion Research in Latin America—*Alberto Castelli*.......... 525
Sondages de l'opinion publique en Amérique latine

United States Methodological Research in Public Opinion Measurement—*Samuel A. Stouffer*.......... 532
La recherche méthodologique aux Etats-Unis sur le sondage de l'opinion publique

Social Security and Health Statistics:
Statistiques des assurances sociales et statistiques sanitaires:

Problems in the Development of Comparable Statistics in the Social Security Programs of Different Countries — *Wilbur J. Cohen*.......... 545
Problèmes du développement de statistiques comparables des programmes de sécurité sociale de divers pays

Problemas del Desarollo de los indices de salud en America Latina —*Manuel de Viado*.......... 557
Problèmes relatifs au développement d'indices sanitaires en Amérique latine Problems of Development of Health Indices in Latin America

Recent Developments in Applications of Statistics to Tuberculosis Control Programs—*Jacob Yerushalmy*.......... 576
Les développements récents dans l'application de la statistique à la lutte antituberculeuse

PART B

DEMOGRAPHIC STATISTICS SECTION

Section de statistique démographique

PAGE

Contributed Papers:
Communications:

The Mortality Experience of the Metropolitan Life Insurance Company, 1911-1946 —*Louis I. Dublin*.......... 595
L'expérience en matière de mortalité de la Metropolitan Life Insurance Company, 1911-1946

La natalité en Italie de 1931 à 1946—*Livio Livi*.......... 601
The Birth Rate in Italy from 1931 to 1946

Mesure des phénomènes démographiques et économiques dans ses rapports avec leur définition—*Alfred Sauvy*.......... 607
Measurement of Demographic and Economic Phenomena in Relation to Their Definition

The First Census of Population Taken in Iceland in 1703 —*Thorsteinn Thorsteinsson*.......... 614
Le premier recensement de la population en Islande en 1703

Outlines of the Demography of the Jewish Population of Palestine —*Rob. Bachi*.......... 624
Aperçu de la démographie de la population juive de la Palestine

Recent Trends in Analysis:
Tendances récentes en matière d'analyse:

The Fertility of Successive Cohorts of Women in the United States —*Pascal K. Whelpton*.......... 631
La fécondité de générations successives de femmes aux Etats-Unis

Analyse des tendances démographiques récentes avec indications spéciales sur les causes et la signification des variations du taux de natalité—*Alfred Sauvy*.......... 660
Analysis of Recent Demographic Trends with Particular Reference to the Causes and Significance of Changes in the Birth Rate

Principios en que se Basa la Clasificacion Estadistica Internacional de Enfermedades, Lesiones, y Causas de Muerte—*Dario Curiel*.......... 672
Principles Underlying the International Statistical Classification of Diseases, Injuries, and Causes of Death
Principes à la base de la classification statistique internationale des maladies, traumatismes et causes de décès

Age Distribution and its Interrelation with the Elements of Natural Increase—*Wilhelm Winkler*.......... 683
Distribution par âge et ses interrelations avec les éléments de l'accroissement naturel

PAGE

Current Problems:
Problèmes actuels:

Current Problems in Morbidity Research—*Haven Emerson*.............. 704
Problèmes actuels dans les recherches sur la morbidité

Evaluation of the Methods of Measuring Net Fertility—*Alfred J. Lotka*.... 715
Appréciation des méthodes pour la mesure de la fécondité nette

Factors of Urban Growth in China—*Ta Chen*.......................... 733
Facteurs d'accroissement de la population urbaine en Chine

Détermination statistique de la consommation incompressible
—*Yves Mainguy*... 748
Statistical Determination of Incompressible Consumption

CONTRIBUTED PAPERS NOT DISCUSSED

Communications présentées pour information

Statistical Relations and Their Inversions (summary)—*Corrado Gini*........ 759
Les relations statistiques et leurs inversions (résumé)

La statistique internationale des forces motrices—*Alessandro A. Molinari*.... 763
International Statistics on Motive Power

Sur le secret en statistique—*Alexandre Dobrovits*....................... 769
About Secrecy in Statistics

Calcul de la différence moyenne—*Nan-ming Liu*....................... 781
Calculation of the Mean Difference

Sur le calcul d'un taux mensuel de mortalité infantile—*Henri Bunle*...... 794
On Calculation of a Monthly Infant Mortality Rate

An Experiment in Collection of Data on Population in Turkey
—*Djelal Aybar*... 805
Expérience dans la réunion de données sur le mouvement de la population en Turquie

Particulars Concerning Turkish Population Census
—*Djelal Aybar*... 818
Renseignements sur les recensements de la population en Turquie
Les régularités dans la proportion des sexes—*Aloys Dolanyi*.............. 826
Regularities in the Proportion of the Sexes

Infant Mortality of the First-born and the Later-born
—*Franco Savorgnan*.. 839
La mortalité infantile chez les premiers-nés et pour les naissances suivantes

Future Development of the Greek Population—*Demetrious Calitsounakis*.. 854
Développement futur de la population Grecque

Methods of Apportioning Seats among the States in the American House of Representatives—*Walter F. Willcox*.............................. 858
Méthodes de répartition des sièges entre les différents Etats à la Chambre des Députés des Etats-Unis

PAGE

Note relative à la statistique intellectuelle—*Vincenzo Castrilli* 868
A Note on Intellectual Statistics

Les problèmes de méthode de la statistique de la production littéraire —*Désiré Elekes* 877
Methodological Problems of the Statistics of Literary Production

Les méthodes d'évaluation des caractères statistiques constituant les guides scientifiques des politiques économique, sociale et fiscale —*Georges-Antoine Hostelet* 886
Methods of Evaluation of the Statistical Characteristics which Constitute Scientific Guides to Economic, Social and Fiscal Policy

The Consequences of Monetary Inflation on the Greek Economy during the War—*Angelos Angelopoulos* 898
Les conséquences de l'inflation monétaire sur l'économie en Grèce pendant la guerre

Some Considerations on the Concept of National Income — *Benedetto Barberi* 906
Quelques considérations sur la notion de revenu national

Methodology in Family Budget Inquiries in the Netherlands Indies —*H. M. J. Hart* 915
Méthodologie d'enquêtes sur les budgets familiaux aux Indes Néerlandaises

The Standard of Living of Native Workers on the Estates in Java —*H. M. J. Hart* 921
Le standard de vie d'ouvriers indigènes dans les entreprises de culture à Java

Essais en Hongrie pour la solution du problème de la statistique sur les budgets de l'Etat *Désiré Laky* 932
A Solution of the Problem of Statistics of the Hungarian National Budget

The National Dividend and Net Production —*Corrado Gini* 942
Dividende national et production nette

OBITUARIES 953
Nécrologies

The following papers, delivered at joint meetings, appear in other volumes as indicated:

Les communications scientifiques suivantes, présentées au cours de séances mixtes sont reproduites dans d'autres volumes ainsi qu'il est indiqué:

Measurement of National Income: (Joint meeting with the Econometric Society)

Mesure du Revenu National:

Recent Experiments in Social Accounting: Flexible and Dynamic Budgets —*J. Tinbergen and J. B. D. Derksen* .... See Vol V.
Expériences récentes en matière de comptabilité nationale: Budgets flexibles et dynamiques

National Income and Industrial Structure—*Simon Kuznets*....See Vol. V
Revenu national et structure industrielle

A Concept of National Income Applicable to Countries Differing Significantly with Respect to the Role of the Government in the Economy—*Arthur Smithies*........................See Vol. V
Une définition de revenu national applicable à des nations dont les gouvernements interviennent différemment dans la vie économique

The Measurement of National Wealth—*G. Findlay Shirras*......See Vol. V
La mesure de la fortune nationale

The Future Role of International Statistical Societies: (Joint meetings with The Inter American Statistical Institute and The Econometric Society)
Le rôle futur des Sociétés Internationales de Statistique:

A Future Role for the International Statistical Institute in International Statistics—*Stuart A. Rice*..................................See Vol. I
Le rôle futur de l'Institut International de Statistique en matière de Statistiques Internationales

A future Role for the Inter American Statistical Institute in International Statistics—*Halbert L. Dunn*...........................See Vol. I
Le rôle futur de l'Institut Inter-Américain de Statistique en matière de Statistiques Internationales

A Future Role for the Econometric Society in International Statistics—*Charles F. Roos*...................................See Vol. I
Le rôle futur de la Société d'Econométrie en matière de Statistiques Internationales

Need for an International Subject Classification for Statistical Materials: (Joint meeting with The Inter American Statistical Institute)
Besoin d'une méthode internationale uniforme de classification des documents statistiques:

Principles and Objectives of an International Classification for Statistical Materials– *Robert Guye*...............................See Vol. IV.
Principes et objectifs d'une classification internationale des documents statistiques

INDEX BY AUTHORS

PAGE

Angélopoulos, Angelos .. 898
Aybar, Djelal .. 805; 818

Bachi, Rob. 624
Barberi, Benedetto 906
Brown, George W. .. 314
Bunle, Henri 794

Calitsounakis, Demetrious 468; 854
Castelli, Alberto 525
Castrilli, Vincenzo .. 868
Chen, Ta 733
Clague, Ewan 178
Cochran, W. G. 40
Cohen, Wilbur J. 545
Cunningham, Ross M. .. 161
Curiel, Dario 672
Curtiss, John H. 351

Darmois, G. 231
Dedrick, Calvert L. .. 17
Delaporte, Pierre J. .. 241
Derksen, J. B. D. .. See Vol. V.
Dobrovits, Alexandre .. 769
Dolanyi, Aloys 826
Dublin, Louis I. 595
Dunn, Halbert L. .. See Vol. I.

Eisenhart, Churchill .. 308
Elekes, Désiré 877
Emerson, Haven 704

Falk, I. S. 212
Fisher, R. A. 434
Flood, Merrill M. 358
Fréchet, Maurice .. 284; 363

Gini, Corrado .. 258; 759; 942
Giusti, Ugo 480
Godfrey, Ernest H. .. 471
Guye, Robert .. See Vol. IV.

Hagood, Margaret Jarman .. 453
Hansen, Morris H. .. 113
Hart, H. M. J. 915; 921
Hauser, Philip M. .. 171

PAGE

Hochstim, Joseph R. .. 129
Hostelet, Georges-Antoine .. 886

Kuznets, Simon .. See Vol. V.

Laky, Désiré 932
Liu, Nan-ming 781
Livi, Livio .. 489; 601
Lotka, Alfred J. 715

Mainguy, Yves 748
Marschak, Jacob 289
Molinari, Alessandro A. 458; 763

Neyman, Jerzy 423

Pabst, Jr., William R. .. 319

Rice, Stuart A. .. See Vol. I.
Roos, Charles F. .. See Vol. I.

Sarle, Charles F. 193
Sauvy, Alfred 607; 660
Savorgnan, Franco .. 839
Shirras, G. Findlay See Vol. V.
Smithies, Arthur .. See Vol. V.
Snedecor, George W. .. 440
Stephan, Frederick F. .. 81
Stoetzel, Jean 518
Stock, J. Stevens 129
Stouffer, Samuel A. .. 532

Taeuber, Conrad 3
Tagliacarne, Guglielmo .. 504
Thorsteinsson, Thorsteinn .. 614
Tinbergen, J. .. See Vol. V.

de Viado, Manuel 557

Wald, Abraham 67
Wallis, W. Allen 331
Whelpton, Pascal K. .. 631
Willcox, Walter F. 858
Winkler, Wilhelm 272; 683
Wold, Herman 302

Yates, Frank 27
Yerushalmy, Jacob .. 576

INTRODUCTION

At former sessions of the International Statistical Institute, the programs included Reports of Commissions established by the Institute, and contributed papers. The program of the Washington session differed substantially from this pattern. The prewar Commissions had not yet resumed their activity and hence had no reports to present. The Bureau and the Arrangements Committee, therefore, selected a number of important subjects of current interest and invited experts to prepare papers on these topics.

This procedure was not entirely new in the Institute. In 1935 the Bureau introduced the category of "invited papers" and such papers were submitted at the Athens and Prague sessions; however, their number was small. For the Washington session, the procedure was applied on a much larger scale.

The 25th session also differed widely from prewar sessions in that it was one of six jointly arranged and interrelated International Statistical Conferences. This fact permitted a number of joint meetings in which two or more organizations participated. The papers submitted at the joint meetings in which the Institute participated are, for the greater part, included in this present volume. The remaining papers will be found in the other four volumes of these Proceedings. As a further consequence the Proceedings of this session are not published separately as Volume XXXI of the *Bulletin of the International Statistical Institute*, but as a part of the Proceedings of the International Statistical Conferences. The material which ordinarily would have been published as parts 2 and 3 of the *Bulletin*, is included in this volume; the material ordinarily appearing in part 1 of the *Bulletin* has been divided between this volume and Volume I of the Proceedings.

G. Goudswaard
Volume Editor

INTRODUCTION

Aux sessions tenues par l'Institut International de Statistique dans le passé, le programme, scientifique était constitué par les Rapports de Commissions établies par l'Institut, et des communications présentées à l'initiative des auteurs. Le programme de la session de Washington s'est écarté largement de cette tradition. Les commissions d'avant-guerre n'avaient pas encore repris leurs activités et, par conséquent, n'avaient pas de rapports à présenter. C'est pourquoi le Bureau et le Comité d'Organisation firent choix de divers sujets importants d'intérêt actuel et invita des experts de préparer une communication sur ces sujets.

Cette procédure n'était pas entièrement nouvelle pour l'Institut. En 1935, le Bureau avait décidé de créer une catégorie de "communications présentées sur l'invitation du Bureau" et de telles communications ont en effet été soumises aux sessions d'Athènes et Prague. Toutefois, leur nombre était limité. Pour la session de Washington, la procédure a été appliquée dans une mesure considérablement plus étendue.

Une autre importante différence avec les sessions d'avant-guerre consistait en ce que la 25e session faisait partie de six Conférences Internationales de Statistiques, organisées conjointement. Ceci permettait de tenir un certain nombre de séances communes de deux ou plusieurs des organisations participantes. Les communications présentées aux séances communes auxquelles participait l'Institut figurent pour la majorité dans le présent volume. On trouvera le reste dans les quatre autres volumes des "Proceedings". Une autre conséquence est que les comptes rendus de cette session ne sont pas publiés séparément pour former le Tome XXXI du Bulletin de l'Institut International de Statistique, mais sont compris dans les Comptes rendus des Conférences Internationales de Statistique. Le matériel qui normalement eut dû constituer les livraisons 2 et 3 du Bulletin est publié dans le présent volume; le matériel ordinairement publié dans la première livraison des Bulletins figure en partie dans le présent volume et pour le reste, dans le premier volume des "Proceedings".

G. Goudswaard
Volume Editor

SPECIAL MEETINGS

WORLD CENSUS OF AGRICULTURE AND POPULATION

*Wednesday, September 17 at 9:30 a.m.**

CHAIRMAN :

Alberto Arco Parró

WORLD CENSUS OF AGRICULTURE

by Conrad Taeuber
Economics and Statistics Division
Food and Agriculture Organization of the United Nations

Never before have governments so urgently needed so many reliable statistics and never have the problems of keeping statistically up to date been greater. In many instances current information and current estimates can no longer be related to the basic data assembled for 1940 or 1935, for these have largely been made obsolete by the course of events since then. In planning their own wartime and postwar programs, governments had an urgent need for reliable objective information about their own resources, and many new statistical services were started. The growing awareness of the interdependence of nations, and the development of international agencies to assist governments in achieving their common purposes, have increased the need for comparability in the statistical data supplied by nations. In an era of frequent international discussion and joint planning by countries, it is essential that their statistical information be as comparable as possible in order to permit the drawing of valid conclusions concerning the relative positions in respect to the matters under examination.

---

* Joint meeting with the Inter American Statistical Institute

In recognition of such needs as these the newly founded Food and Agriculture Organization of the United Nations was asked by its member governments to assume leadership in the development of a broad cooperative enterprise which has come to be called the 1950 World Census of Agriculture. Under the program, each nation is asked to conduct a census of agriculture in 1950 or proximate years. Each participating country will conduct the census according to its own practices and in the light of its own administrative situation, but there is to be prior consultation among nations to assure that there will be comparability in the concepts and definitions used. A minimum set of tabulation plans is to be agreed upon in advance, and thus each country in planning its work can make provision for tabulations which will be comparable to those of other countries. The FAO is expected to combine the tables supplied by the several countries and in that way develop an inventory of the world's agriculture.

This cooperative activity among countries assumes that through cooperation countries can enhance the value of their own efforts in respect to their census programs, and that no one country and no one organization possesses a monopoly of the skills and knowledge necessary to conduct a successful census under the wide variety of economic, social and administrative circumstances which are found in the countries of the world. Through joint study of the technical questions that arise in all stages of census undertaking, and through sharing of the procedures that have been developed, there will be a growing use of the best methods which the technicians of the several countries can devise. The role of the international agency in this undertaking is primarily one of facilitating the free exchange of ideas, information and knowledge among the technicians and administrations of all countries, and not one of attempting to impose a rigid pattern of operations on all countries.

WHY A CENSUS IS NEEDED

The need of a government and its citizens for objective, accurate, and timely information about its agricultural resources and its agricultural plant is so widely recognized that it is not necessary to belabor the point here. When the International Institute of Agriculture began its activities in behalf of a 1930 World Census of Agriculture it found that 37 countries had taken an agricultural census in the first 25 years of the present century. Today, periodic censuses of agriculture are provided for by law in a number of countries, some at ten year intervals and some at five year intervals. These provide a comprehensive picture of the agriculture and of the agricultural resources of the country. In addition, a considerable number of countries conduct annual enumerations on a limited number of items such as crop acreages, acreage sown, acreage harvested, livestock numbers, etc. Where a crop or a branch of livestock production assumes particular economic importance for the country, there may be enumerations at appropriate intervals throughout the season, in order to provide accurate and up-to-date information at all times. In many countries which do not have census experience or a census tradition, there have been efforts to arrive at estimates in lieu of census information, in order that the government and its citizens should have at least a minimum of data on which to base their planning and activities. Some governments have initiated a census program by carrying out enumerations in a district or province at a time in order to gain experience before doing the work on a national basis.

An agricultural census is an inventory of the nation's agricultural resources and plant, of producing areas, of livestock, of the people who work in agriculture, of the conditions under which the land and other agricultural resources are made available to agricultural workers, of animal and mechanical power and of equipment and of other important facts about agriculture which influence the production of food and raw materials and the prices that producers received and consumers pay.

The taking of an inventory is a well recognized practice in many types of private and public endeavor. When combined with information concerning current activities, it provides an invaluable aid in evaluating the successes or failures of the past. It reveals the present position, and provides a guide to planning for the future. Countries also need a periodic stock taking. The inventory of their agriculture which is supplied by a census gives the country and its citizens (especially the agriculturalists and their representatives) the information which is indispensable as a guide to programs for future production and development. It serves as a guide to marketing and distribution programs by showing the distribution of production, and providing a basis for comparisons with facilities for marketing and transportation. Alert businessmen, whether operating privately or as the representatives of cooperative or publicly owned enterprises, have found such data of great value in planning their own operations. In the same way, manufacturers and distributors of manufactured products have found the data on the distribution of a country's agricultural resources and production a guide to their own production and distribution programs. The agricultural producers themselves benefit through the more efficient planning for agricultural production and their own welfare which is possible when full information is available. An important direct benefit grows out of the reduction of harmful speculation in the prices of agricultural commodities, for that type of speculation thrives on uncertainty and lack of information by the public.

Census data are unsurpassed aids to governments in the development of intelligent action and legislation for the promotion of agriculture and the welfare of rural people. The census provides data not only for the entire country but also for the smaller geographic areas, which may have widely differing conditions and may require special consideration and treatment. A census moreover brings together a body of information concerning the individual agricultural enterprises of the country, and thus provides the basis on which legislators, administrators, research workers, scholars, journalists, and others can study relationships among the factors affecting the country's agriculture in a way which makes for much more intelligent understanding of conditions and problems that would be possible without such sources. That this is a possibility which has not always been fully utilized in the past is regrettable. The development of newer techniques for the handling of mass data, and a growing collaboration between census technicians and the consumers of their data should make for greater realization of the possibilities of such use than has previously been the case.

When countries join with each other to promote the increased production of food and agricultural products throughout the world, the improvement of distribution of such products, the improvement of nutritional levels and the increases of the welfare of their people, they find that facts about their agriculture are indispensable. They need the facts concerning themselves in order to evaluate the progress which they themselves are making, but they also need the facts concerning the other countries with which they are cooperating in order properly to evaluate the effects of trends and activities in those other countries upon their own people and upon the entire international community. To carry out such an activity there is need for the free interchange of information, and a recognition of the fact that there is no lasting advantage to the individual country in keeping the others in ignorance of the true state of affairs.

THE PROGRAM FOR 1950

Realizing the importance of a joint census undertaking at an early postwar date, the delegates to the first Conference of the Food and Agriculture Organization in 1945 instructed the Director-General to study the problems involved in promoting a world census of agriculture in 1950 and to report at the subsequent conference. At the 1946 Conference it was agreed that the work on the 1950 World Census of Agriculture should become one of the early activities of the Organization, and accordingly steps have been taken to get under way with this cooperative endeavor.

The first step was to prepare a minimum program which might be proposed for consideration by all governments, including the responsible authorities in non-self governing areas. The international *ad hoc* group which was convened for this purpose did not believe that it would be feasible to attempt to draw up a complete questionnaire to be proposed to governments. Instead it preferred to develop a list of items for which information is to be secured, and to develop a set of definitions and tabulation plans to accompany this list. In developing the minimum list the needs of countries which do not have extensive census experience were particularly kept in mind, and the list was confined to those items which are needed from all countries. It was not intended that this minimum list would satisfy the needs of all countries—in fact it may actually satisfy no one country's needs completely. Adaptation to meet local conditions and needs, further subdivision of the items included in the minimum list and additions to that list are not only assumed, but are actively encouraged. In the field of crops, the minimum list was confined to those crops for which world-wide totals were deemed to be desirable because of the importance of the crops to nutrition or to the international trade in agricultural products. Each country will need to supplement this crop list in order to provide information on all of the crops which are of importance to it.

In addition to the minimum list, another list, providing for greater detail for some subjects and for additional subjects has been prepared. Its purpose is to provide a basis for securing comparability for the items to be included in the censuses of countries which will wish to do a more comprehensive census than is provided for in the minimum list. The two lists together provide definitions for a large number of items which are ordinarily included in census activities, suggest uniform classifications for livestock, and many other items which have given trouble in international comparisons in the past. In developing these lists full use has been made of the experience gained in the similar undertakings for 1930 and 1940.

These lists of items have been submitted to governments as provisional and criticisms and suggestions have been invited. They were also submitted to many statisticians and other technicians in all parts of the world for their criticisms. In March 1947, the European governments were invited to send their representatives to Rome to consider these proposals and other aspects of the Census program in detail. Representatives from 18 countries were present at that meeting. Representatives of the American governments have just completed a similar review of the proposed list of items as part of the work of IASI in relation to the 1950 Census of the Americas. Similar meetings are proposed in the Middle East and in Southeastern Asia later in 1947.

Discussions have been initiated with representatives of Colonial offices to consider ways in which the program might be adapted to the needs and conditions of many of the dependent areas. After these discussions have been completed, a definitive list of items is to be submitted to governments. When this list is submitted it will have had the benefit of the views and experience of statisticians in a large number of countries.

The question of developing a minimum or model questionnaire was considered at some length and was finally decided in the negative. It was considered that a questionnaire is a tool for securing information which must be adapted to the collection procedure appropriate to each country. This is related to past experience with censuses and to economic, social, geographical, and political conditions in the country. A questionnaire which is carried to the agriculturalist by an enumerator specifically employed to collect the information will be quite different from one that is taken to the agriculturalist by the enumerator and left to be filled in and mailed at a later date, or one that is to be used when the enumerator's function is primarily that of assisting the agriculturalist complete a return which he has largely filled out by himself. A different form is required for use in a situation in which it is obligatory for each agriculturalist to come to a specified place, such as the headquarters of the village government, to supply the required information in the presence of witnesses. The degree of literacy in the population has a profound effect on the structure and content of the questionnaire. That is true also of the degree to which the agriculturalists are subject to governmental controls in respect to their agricultural operations. The predominant forms of tenure and the character of agricultural holdings significantly affect the development of a questionnaire. Perhaps most important of all is the realization that an effective questionnaire must be so developed that there is no misunderstanding on the part of the respondent concerning the information which he is to supply. That means that the questions must be framed so that they will have the same meaning to the respondent and to the statistical office. This is not only a matter of language but is essentially a matter of an intimate knowledge of the conditions under which agriculture is carried on in the different areas of a country. It is essential that a questionnaire should be developed in relation to the tabulating procedures to be followed. These in turn depend on the administrative structure of the governmental and on the type of tabulating equipment which is available.

Although no uniform questionnaire appeared to be feasible, it was deemed feasible to develop minimum standards to which questionnaires should correspond. In order to assist in the maintenance of comparability in the returns which will be secured in the several countries, provision has been made for consultation between representatives of FAO and of the statistical services of the countries. In each of the major regions of the world a representative of FAO will be available to assist any government in the development of its census program and materials with a view to insuring comparability, while maintaining the maximum flexibility in adapting to local conditions. It is planned to summarize and publish information concerning practices developed in countries which already have census experience and to set forth in some detail practices which have been found to be effective under different conditions. A minimum set of table outlines is in process of revision for transmission to governments so that they may plan their tabulations in such a way that comparability in the final tables will be assured. FAO stands ready to assist any government on request in finding trained technical personnel which may be seconded to the government to assist in the census.

Recognizing that adequately trained personnel is vital to the successful conduct of any census program, particular attention has been given to means whereby training in census procedures will become more readily available. In a number of countries with well developed statistical services, or with well equipped training facilities, steps are being taken to initiate training which combines practical experience in census procedures with theoretical training in statistics. Arrangements have not yet been completed but it is hoped that a number of countries in different parts of the world will make their facilities available to nearby countries, and that the necessary arrangements can be made so that trainees may benefit from such facilities. In this connection, full advantage is being taken of the training facilities being made available by the United States to persons from other American countries. Countries which have not had recent census experience are particularly in need of such assistance. But the need is not confined to them alone. Indications have come from a number of countries that the war-engendered shortages of technically trained personnel have also affected their statistical services and that there is an immediate need for training programs if adequate personnel is to be available in time for the census in 1950. A member of the FAO staff is now devoting full time to this problem, and the prospects are encouraging. Consideration is being given also to the possibility of establishing short term intensive training schools in a number of areas.

DEFINITIONS AND PROBLEMS

In the 1950 World Census of Agriculture it is intended that each government will obtain accurate and comparable information on its agriculture including

(a) areas under crops and numbers of livestock

(b) the number of agricultural holdings and their principal characteristics, such as size, form of tenure, utilization of the area, utilization of labor, implements and mechanical power, etc.

(c) the number and characteristics of people who secure their livelihood from agriculture, and the extent to which they also engage in other occupations, especially in forestry and fisheries, and

(d) a record of its production of all important agricultural products.

It is proposed that in every country the 1950 Census of Agriculture shall cover all types of agricultural holdings, including those whose product is used primarily for consumption by the holder and his family. It is recommended that tribal and other groups which are outside the normal marketing system be included even though for them it may be necessary to resort to means other than individual enumeration of holdings. In this respect the effort for 1950 covers a broader field than that of the earlier World Census of Agriculture. Then the emphasis was somewhat more on those parts of agricultural production which entered into commercial channels. In broadening the scope of the census, recognition is given to the purpose of governments to take appropriate steps to improve the nutrition of all their people and to improve the welfare of rural populations, which was accepted when they became members of the Food and Agriculture Organization.

The program as submitted to governments avoids setting a minimum area or minimum output of the holdings to make them eligible for inclusion in the enumeration. Instead governments are urged to make the minimum limit as low as possible and to inform FAO as soon as possible of the limits adopted. It is expected that agreement will be secured on the size groups according to which tabulations are to be presented, and thus the materials can be presented in comparable form.

The 1950 World Census of Agriculture is in principle conceived of as a direct enumeration of individual holdings in each country. It is recognized that in some areas, especially those where agriculture is carried on by primitive people, it may not be practicable to secure a report from each individual producer and hence it will necessary to prepare estimates for the entire community, tribe etc. In such cases it is recommended that a report similar to that secured for individual holdings should be obtained for the appropriate unit. In such cases the area covered should be limited to the land under cultivation and the pasture lands actually used. In all cases in which such estimates are to be made it is especially important that the local organizations or officers responsible for the estimates receive clear and precise instructions indicating the records or other sources to be used and the methods by which the estimates are to be made. In such cases avoiding omission and avoiding duplication of some areas or items becomes especially important.

In so far as possible it is intended that the 1950 World Census of Agriculture should relate to the crops harvested in the northern hemisphere in the calendar year 1949 and in the southern hemisphere in the year from July 1, 1949 to June 30, 1950. It is recognized that in some countries the statutes prescribe another year, and in some other countries there may be unavoidable reasons for having the data relate to another year.

Although it is desired to secure information concerning the crops harvested, it is not deemed feasible to recommend as a general practice that agriculturalists be asked to report directly their production of the several crops. In many countries the methods of harvesting, the threshing practices, and utilization practices make it difficult for the holder to report with any degree of accuracy the total production of the crop realized on his holding. In addition, there are some areas in which resistance to reporting production is likely to be encountered because respondents are not fully convinced that the legal provisions for holding the returns confidential will actually be observed. Therefore, chief attention is to be given to areas harvested and it is strongly urged that every government take the census as soon as possible after the main harvest season has ended for the year (usually when the bulk of the row crops have matured or have been harvested) and before general planting operations have started for a new crop season. Some countries have, by custom or law, taken censuses at another period in the year. In a number of countries enumerations are customarily made of areas sown instead of areas harvested and for some of these it has been ascertained that there is ordinarily very little difference between area sown and area harvested. However, in all instances in which a country cannot enumerate area harvested, it is urged to supply estimates of the differences between the area sown to different crops and the area actually harvested.

The definition of an agricultural holding or farm is essentially the same as that used in the 1930 and 1940 censuses. A holding according to this definition includes all land used wholly or partly for agricultural production that is operated, directed, or managed by one person alone or with the assistance of others, without regard to title, size or location. It may consist of two or more parcels of lands, even if widely separated, provided that they form a part of the same technical and economic unit. Enterprises not including any cultivated land but producing livestock or livestock products are to be considered as farms or agricultural holdings, regardless of their location. Essentially, in order to be classified as an agricultural holding, an enterprise must produce agricultural products and must be a unit in the technical and economic sense.

Such a generalized definition leaves much to be worked out in the individual countries. It is clear for the great bulk of agricultural holdings in all parts of the world including those under individual proprietorship or leased by an individual, and those which are operated under some pattern of cooperative, communal, or public ownership or management. There are numerous problems still to be worked out in this connection, for the variations in tenure relationships are infinite. Some types of plantation and multiple holdings—such as those in parts of the southern United States—do not readily lend themselves to clear-cut classification. There are many degrees of economic and technical responsibility between that of the laborer who assumes no responsibility and that of the proprietor who operates his own holding and assumes full responsibility. These situations will require careful study in order to arrive at consistent, meaningful, and comparable classifications. In some cases it may be necessary to request the government to secure the information in such a way that alternative methods of classification may be utilized.

The problem of securing information on the production of livestock and poultry products has been the subject of extended study. The nutritive value of such products, particularly meat, milk, and eggs, makes it especially urgent that information about them be secured. The fact that these products, particularly milk and eggs, are "harvested" daily and that the yield per unit is subject to wide seasonal variations makes it particularly difficult to secure useful information for such products. For these items it is proposed to secure through the census enumeration information for only one day, *i.e.*, the day preceding the enumeration. This is a figure which the holder should be able to report without difficulty, and in most instances the enumerator collecting the information could verify the holder's report, should that appear necessary.

This is at best a partial answer. Such figures by themselves cannot be used for making comparisons of rates of production of animal products or of total production in the several countries. When supplemented with other information, they can, however, be used as a basis for estimating total production during a year or parts of a year. Because production varies from season to season, governments are urged to prepare estimates of seasonal changes in production or estimates of production for a week, a month, and wherever possible, a year.

The treatment proposed for data on the production of crops and livestock products illustrates one basic element in the approach. That is to view a census as an important element in a program of agricultural statistics, but not as the only part of such a program. In order to make data on crop acreages useful they must be supplemented by estimates of yields. Governments which have made such estimates in the past are generally agreed that yield estimates do not lend themselves readily to the techniques of census enumerations. Governments that have not in the past made current estimates of production will find the census data on crop acreages indispensable in order to interpret the yield estimates. Likewise, the data on livestock products proposed in the census program will be of very limited value unless there also are estimates for parts of the production cycle other than the one included by the date of the enumeration. On the other hand, current estimates of the yield of livestock products would benefit substantially from the objective and comprehensive data supplied by an enumeration.

The value of the data collected in a census is enhanced if at the same time related information is secured, and if the several data collecting activities are closely coordinated. Because of the close relationship among agriculture, forestry, and fisheries, it is suggested that wherever feasible, censuses or surveys of certain aspects of forest industries and fisheries be undertaken concurrently with the 1950 Census of Agriculture. It is also proposed that the data on use of the land which is included in agricultural holdings be supplemented by estimates of the total land area of the country and of the major uses of the land not included in agricultural holdings. The information to be secured as part of the census of agriculture is necessarily limited to the land included in agricultural holdings. If the other information could be secured at the same time, both sets of data could be more readily related to each other, and one check would be provided on the completeness of the census enumeration.

Information on the number and characteristics of people who secure their livelihood from agriculture and the extent to which they also engage in other forms of gainful activity is regarded as primarily a matter for a census of population, and it is urged that wherever possible, the census of agriculture be coordinated with a census of population and the data from the one be directly related to the data from the other. In cases in which that is not possible it is proposed that a brief section on the agricultural population be included to provide a minimum amount of information on age and sex distribution and on the degree of participation in agricultural activity.

Because of the possibility of enhancing reliability, saving of funds, and speeding the publication of results, there is growing interest in the role of sampling in a census program. Sampling has been used in connection with census activities for four major purposes: (a) to secure additional information which could not be secured from all persons included in a complete enumeration, (b) to test the reliability of the census, (c) to reduce costs and speed up tabulations, especially the more complex tabulations, and (d) in lieu of a complete enumeration. Sampling methods have also been found effective in the control of various operations of a census, enumeration, coding, editing, and card punching. How sampling may best be used as part of the census program will depend on the uses to which the data are to be put. Since a sampling program must be adapted to the purposes to be served and the situation in which it is to be applied, no one sampling method for general use has been proposed. It is proposed, however, that the rapid advances in the application of sampling methods to enumerative procedures should be utilized in connection with the 1950 World Census of Agriculture.

This report is an interim report on an activity that is now under way. As of September 15, 1947, thirty-seven countries had stated that they expected to participate actively in the 1950 World Census of Agriculture, and many others have indicated an interest. A definitive proposal will be submitted to governments as soon as the proposed discussion with technical representatives of countries have been completed. The more detailed process of working out the necessary national adaptations to the over-all program is under way and will continue. A training program is expected to be well under way by midwinter. Materials summarizing the experience of countries that have taken censuses in the past are being prepared and will be made generally available. The staff of FAO stands ready to assist governments which requests assistance in the census undertaking, and it has been assured the assistance of countries that have well developed statistical programs. FAO plans to publish international tables summarizing the results as quickly as they become available. Since this is a cooperative undertaking involving many governments and many persons, it is not possible at this time to state definitely what the outcome will be. This is the third time that a World Census of Agriculture has been discussed before the International Statistical Institute. We look forward to the time when we shall be able to report the successful completion of so ambitious an undertaking.

Résumé

La reconnaissance de plus en plus général de l'interdépendance des nations, et la création d'organisation internationales pour aider les gouvernements à atteindre leurs objectifs communs, ont fait grandir la nécessité de statistiques comparables. Reconnaissant cette nécessité, les Etats Membres de la FAO (Organisation des Nations Unies pour l'Alimentation et l'Agriculture) lui ont demandé d'assumer un rôle dirigeant dans le recensement mondial de l'agriculture en 1950.

Le rôle de la FAO dans cette entreprise est principalement de faciliter, entre les techniciens et les administrateurs des divers pays, des échanges de connaissances, d'idées, et d'informations concernant les techniques et les méthodes employées, plutôt que de tenter d'imposer à tous les pays l'usage d'un questionnaire uniforme et rigide.

On a tout d'abord convoqué un groupe international *ad hoc* pour préparer un programme minimum à soumettre à tous les gouvernements, y compris les autorités responsables des territoires sous tutelle. Le programme minimum comprend une liste de matières pour lesquelles il est des plus urgent d'obtenir des statistiques internationales, ainsi que des notes explicatives et les définitions proposées. Une seconde liste, plus complète, comprend outre plus de détails sur certaines questions, des matières supplémentaires sur lesquelles il est désirable d'obtenir des renseignements.

Après que le programme fut soumis aux gouvernements, des conférences eurent lieu à Rome (en mars 1947) et à Washington (en septembre 1947), au cours desquelles ces propositions et d'autres aspects du programme furent étudiés en détail par les représentants des gouvernements européens et américaines.

Etant donné la nécessité de disposer d'un personnel bien entraîné, une attention toute spéciale a été consacrée aux moyens par lesquels l'entraînement dans les méthodes de recensement pourraient être plus largement répandus. Dans certains pays où les services de statistique sont bien organisés et outillés, on s'occupe de préparer des cours d'entraînement comprenant tant l'expérience pratique des méthodes de recensement que l'enseignement théorique de la statistique. On envisage aussi la possibilité d'établir, dans certaines régions, des centres d'entraînement intensif de courte durée.

On vise à ce que chaque gouvernement obtienne pour le recensement mondial de l'agriculture en 1950 des renseignements comparables et exacts en ce qui concerne l'agriculture de son pays, comprenant tous

les genres d'exploitations, en complétant les données recueillies lors du dénombrement complet, par des évaluations et des sondages.

Au 1er septembre 1947, 37 pays avaient fait savoir qu'ils participeront au recensement de 1950, et plusieurs autres avaient montré un certain intérêt dans l'entreprise. Les travaux d'adaptation des plans aux diverses nécessités d'ordre national ainsi que le programme d'entraînement statistique sont en bonne voie.

La FAO envisage la publication de tableaux internationaux résumant les résultats du recensement au fur et à mesure qu'ils seront disponible.

THE 1950 CENSUS OF THE AMERICAS

by Calvert L. Dedrick
Coordinator, International Statistics
Bureau of the Census (United States)

It is appropriate to note at the beginning of this brief report on the plans for and problems of the 1950 Census of the Americas that just fifty years ago the International Statistical Institute, at its St. Petersburg session in 1897, considered recommendations which could be made for a world census of population to be taken in the year 1900. The pioneering efforts of the ISI in attempting to secure international uniformity in methods of collecting and presenting census data should be recognized by everyone interested in plans for internationally comparable censuses.

Statisticians from several of the American nations were present at sessions of the ISI which discussed census plans, and they were instrumental in securing the passage by the Conferences of American States in 1910 and 1923 of resolutions recommending that the participating governments take decennial censuses of population and exchange technical information as well as statistics. Although these early recommendations as to standards and international cooperation were helpful, they were not entirely successful because they lacked leadership and an organization devoted exclusively to this purpose. It appears that no really effective steps were taken to coordinate the censuses of the American nations in 1930 or 1940.

The genesis of the present activities for a 1950 census of the Americas is not found in these early recommendations and resolutions, but in the fortunate combinations of a *leader* who can make ideals a living reality, an *organization* devoted exclusively to inter-American statistics, and an *inter-American congress* considering demographic problems.

The leader is Dr. Alberto Arca Parró who, as chairman of the Demographic Statistics Committee of the Inter American Statistical Institute (the organization), and Director General of Statistics in Peru, wrote to the other Director Generals of the Americas on June 7, 1943, suggesting the great desirability of a population census of all the American nations in 1950. In this letter Dr. Arca suggested the first practical steps toward such a census as: (a) the exchange of census documents and experience; (b) the technical determination of the type and minimum content of a decennial census form which could be used in each country without prejudice to its amplification to meet national needs; and (c) the preparation of national laws providing for such a census, and for necessary financing. The response to this suggestion was immediate and favorable.

The Inter-American Demographic Congress, which met in Mexico City in October 1943, considered and approved Dr. Arca's proposal. It formally requested the Inter American Statistical Institute to take such steps as might be necessary for an inter-American population census in 1950. A project for a census of the American nations was immediately incorporated in the program of the Institute and, at the first meeting of the Executive Committee, in January 1946, a new committee was created for this purpose.

Additional endorsements of this project were received from the Pan American Consultation of Geography and Cartography (Rio de Janeiro, Sept., 1944), the Inter American Conference on Agriculture (Caracas, July 1945), the Inter American Defense Board (Washington, August 1946), and the Economic and Social Council of the United Nations (March 1947) after recommendations by both the Statistical and Population Commissions.

Between September 2 and 16, 1947, nearly 100 men and women representing all of the 22 American Nations participated in round table discussions and plenary meetings of the First Session of the Committee on the 1950 Census of the Americas. The Committee is composed of one technical representative of the general statistical agency or, in some cases, the population census agency, of each of the American Nations, and of observers from three international organizations : the United Nations; the Food and Agriculture Organization; and the Pan American Union. This Committee was supported by nearly 70 technicians from the countries involved.

Time does not permit an adequate presentation of all phases of the Committee's work, including the working papers considered or future plans of the 1950 Census of the Americas. Documentation of the program will be available in the near future. This paper, therefore, will be limited to the presentation of a selection of the principal actions of the Committee and a brief outline of some of the technical and administrative problems which remain to be solved.

The 1950 Census of the Americas is really not *one* census but twenty-two. Each country of the Western Hemisphere is now planning to take a census of population during, or within six months before or after, 1950. As always, these censuses will be taken primarily to provide data for the nation itself but with the new objective of using, insofar as possible, a uniform list of topics and classifications which will provide date on an internationally comparable basis.

The Committee has recognized that if there are to be international standards for statistics which will be used in all countries, they must be very practicable standards which are designed to satisfy the needs of these countries. The Committee has demonstrated, in the opinion of the writer, that the best way to arrive at such standards is to sit down around a conference table and, after adequate discussion of the mutual and special needs of national and international organizations, to arrive at provisional items which can be adopted in principle. The next step, of course, is to test these standards and, in future meetings, to adjust and readjust them until they fit insofar as possible both national and international needs.

One of the fundamental objectives of the first session of the Committee on the 1950 Census of the Americas has been the adoption of uniform items and definitions for population census purposes, and uniform methods and procedures which would provide comparability of results among the 22 censuses were also considered.

During this first session the Committee has reached a substantial agreement as to certain items which should be included in the population census program of the American Nations. For example, the following items have been included in the inter-American minimum population census program without much difficulty:

1. Sex
2. Age
3. Place of birth (For the native born, the State or other major territorial division. For the foreign born, the country they report.)
4. Naturalization (For the foreign born, indicate if the person is naturalized in the country in which the census is taken.)
5. Literacy (Does this person know how to read and write? Yes, or No.)

In addition to these, a number of items have been included in the minimum census program subject to further study as to the exact definitions and classifications to be used. Such studies will be conducted under the auspices of the coordinating board of the Committee and the results will be considered in the next session of the Committee. All of the studies on these items will take into account consultations with national and international organizations. Items included in the minimum program but subject to further study are the following:

1. Relationship to head of the family (principal problem—the "census family")
2. Marital status (principal problem—common law marriages)
3. Instructional level (principal problem—an international measure of the amount of schooling)
4. Mother tongue
5. Gainfully occupied population (principal problem—definition of groups to be included, applicable throughout the hemisphere ;

and adequate international classifications of the following items concerning each person included as belonging to the gainfully occupied group):

a. Individual occupation
b. Industry
c. Class of worker

The Committee did not include in the minimum census program, at this first session, the following subjects for which definitions and classification standards are to be studied. On the other hand, because of their outstanding importance, they are considered as possible candidates for later inclusion in the minimum program provided adequate international definitions and classifications can be developed:

1. Nationality
2. Income from wages and salaries
3. Fertility
4. Cultural characteristics (of the indigenous populations, for regional use)
5. Data relating to dwelling units for possible inclusion in the population census questionnaire in the event that a separate census of housing cannot be taken.
6. Urban and rural population (especially the definition of a "place")
7. Agriculture population, *i.e.*, total persons economically dependent on agriculture.

The Committee is also recommending for the minimum inter-American population program the securing of a *de facto* total population of the country as a whole even in those countries which by law are required to take a *de jure* census. It is also recommending that the enumeration of population be conducted as rapidly as possible, preferably in a single day, and if possible, in not more than one week in rural areas. As to method of enumeration, the Committee has indicated a preference for enumeration by house-to-house canvassers, though without prejudice to the use of forms made out by the head of the household in those cases where this is feasible.

The Committee on the 1950 Census of the Americas assigned to one of its principal subcommittees the problem of aiding the Food and Agriculture Organization of the United Nations with the problem of agricultural censuses in the American Nations which will be taken in or about 1950 as part of the World Census of Agriculture. The Committee is interested both in securing agriculture censuses in this hemisphere and in the coordination of these censuses with population inquiries as part of the 1950 Census of the Americas.

The Committee has also undertaken as future work the development of international standards in economic censuses such as commerce, industry, transportation, etc. The tremendous need for data of this type is amply recognized, and studies will be made by the Committee in cooperation with the national statistical offices of the American nations and its parent organization, the Inter American Statistical Institute, so that as much information as possible can be considered at the next meeting of the Committee in 1948. It is of considerable interest, I believe, that the Committee has indicated that one of the objectives of these economic censuses is the securing of data which will facilitate the calculation of national income estimates.

Time does not permit the review of all the other actions of the Committee, but a brief listing of some of the points on which agreements have been reached includes: census legislation; the organization of census offices; cartographic materials for census purposes; tabulation procedures; a published guide to census procedures and analysis; the financing of the Committee's work; provision for future sessions of the Committee; regional Institutes for training census personnel; and relationships with the United Nations and other international organizations. On the last point it is significant that the Committee has instructed its governing board to consult with the international organizations interested in the Census of the Americas and to use international definitions and classifications wherever possible.

The Committee expressed its appreciation to the Economic and Social Council and the Population and Statistical Commissions of the United Nations for their interest in and approval of the Census of the Americas. The Committee's Chairman is instructed to keep the Secretary General of the United Nations informed as to resolutions and other matters resulting from activities of the Committee and, in return, the Secretary General of the United Nations is requested to keep the Committee informed as to the development of uniform standards and international census plans of the Economic and Social Council and its commissions.

Great progress has been made in making the 1950 Census of the Americas a reality. But a tremendous amount of work remains to be done both nationally and internationally. Quite apart from the technical problem of attaining comparability of information among the censuses of the Americas, and not subject to the same type of scientific research and agreement, are the very grave administrative problems which must be faced if all of the American nations are to take a census in or about 1950.

Without attempting to make an exhaustive list of such administrative problems, the writer calls attention to the following:

1. Several of the American nations are almost completely lacking in census traditions, *i.e.*, in the public acceptance of and cooperation

in periodic population censuses. For example, one country has not had a census since the period of its independence; two countries, though they have had enumerations, have not published tables even on such basic population facts as sex and age. Two more have been without census of population since 1908 or earlier. A number of other countries as well are not conditioned to regular enumerations. Without this census tradition the public has scant interest at best and, at worst, is openly opposed to the enactment of adequate census laws and the provision of sufficient funds to make periodic censuses possible. Under these circumstances there is need for courageous leadership on the part of public administrators who realize the tremendous need for population data.

2. A number of countries have no continuing census organization. This is one of the most serious defects in the present statistical organization of the American nations, and one of the major factors contributing to the cost of population censuses in many countries. Fortunately, there is a growing realization of the importance of maintaining at least a cadre of trained personnel within the national statistical organization during the interim between decennial enumerations.

3. A third administrative problem and one quite closely related to, though independent of, the lack of census tradition and lack of organization, is the lack of sufficient finances for the census program as a whole. The American nations have been fairly liberal in providing funds for the enumerative phase of the census. But these funds have often been made available so late that there were inadequate expenditures on preparatory work. Too often the interest of high public officials and of the public has lagged immediately after the enumeration with the result that insufficient funds were made available for the adequate tabulation and publication of data. It should be one of the objectives of each of the American nations, in preparing for its census in or about 1950, to budget its census funds in such a manner that money is available for every phase of the census.

4. A fourth administrative problem of considerable importance is that of integrating in advance the results of the censuses of population and agriculture with current statistical data in many fields such as vital statistics, education, criminal-judicial statistics, cost of living, agricultural estimates, etc. To be successful, this integration must be made in the very early stages of census planning. The exact manner of integrating the current statistical programs of many agencies with the major project of a decennial census, is a matter which must be worked out in each country in accordance with its own laws, traditions and existing organizations.

In conclusion, I wish to quote from a speech on inter-American censuses given by S. N. D. North, former director of the U. S. Census, at the First Pan American Scientific Congress, Santiago, Chile, December 25, 1908, to January 5, 1909:

"General Francis A. Walker, who was superintendent of the 10th and 11th Censuses of the United States,..once remarked that 'the people of the United States are well able to pay for the very best census they can get;' and the people have proved each decade that he was right.

"If I may be permitted to paraphrase his remark, I will conclude by saying that the people of no South American country can afford *not* to pay for the very best census they can get, periodically taken, at least once every ten years, covering as many lines of national activity as possible, and taken in accordance with a uniform plan."[1]

Résumé

Du 2 au 16 septembre 1947, une centaine de statisticiens représentant les 22 nations américaines ont pris part à la première session du Comité du Recensement des pays de l'Amérique en 1950, organisée par l'Institut Inter-Américain de Statistique. L'un des objectifs principaux de cette session était l'adoption de définitions uniformes, et, dans la mesure du possible, de méthodes et procédures uniformisées pour le recensement de la population.

Le Comité a abouti à un accord de principe sur certains points qui devront figurer sur le programme minimum du recensement de la population. Les points suivants furent adoptés sans grande difficulté: sexe, âge, lieu de naissance, nationalité, et analphabétisme. En outre, le programme minimum devra contenir les questions suivantes qui, toutes, font l'object d'une étude plus approfondie quant à leurs définitions et classifications: parenté avec le chef de famille, état matrimonial, degré d'instruction, langue maternelle et, pour la population active, la profession individuelle, l'industrie où on l'est occupé et la position dans la profession. Plusieurs autres questions furent réservées pour un examen ultérieur. Le Comité a décidé également de s'occuper de l'uniformité internationale dans les recensements économiques

---

[1] North, S. N. C., "Uniformity and Cooperation in Census Methods of the Republics of the American Continent," *Amer. Stat. Ass'n. Journal*, No. 84 (December 1908), p. 295 *ff*.

(commerce, industrie, moyens de transport, etc.) des pays de l'Amérique.

Un grand progrés a été réalisé en ce qui concerne le recensement des pays de l'Amérique en 1950. Toutefois, ce progrès a été fait en grande partie dans le domaine technique, et si les 22 nations américaines doivent procéder à un recensement en 1950, plusieurs problèms très importants d'ordre administratif devront encore être réglés. Ce sont: la création d'une opinion publique favorable dans les pays dépourvus de tradition dans le domaine des recensements périodiques, la création de Bureaux de Recensement permanents, la formation du personnel nécessaire, l'obtention de crédits et l'intégration du programme de recensement dans les statistiques courantes de chaque pays.

Discussion

Mr. Davis felt that Mr. Taeuber had overstated the value of an agricultural census. He recommended that previous census experiences be analyzed by the present planners. Mr. Davis urged limitation of scope rather than incomplete or imperfect coverage within a broader scope.

Mr. Davis urged prompt publication of results, with exaplanations of the data to prevent their misuse.

Mrs. Czerniewska discussed the difficulties which will be involved in the taking of a Census of Agriculture on the European continent. In the first stage, methodological and technical preparation of the census, most European countries have tradition and experience. The decision to incur the expense of a census, however, will be a difficult one for the European nations to make.

The carrying out of the enumeration, said Mrs. Czerniewska, will present the greatest difficulty; because of wartime requisitions, quotas, etc., the agricultural population will have a hostile attitude toward the enumerator. Mrs. Czerniewska commented that means must be devised to allow for the inclusion of some lands which are operated by communities.

The handling of results, finally, is of great importance to the success of the census. Mrs. Czerniewska concluded that it is most important that the FAO reach an agreement with the participating countries, which have already approved the minimum program, with respect to the order of handling and a timetable for delivery of results.

Miss Miró commented on the three problems which the Committee on the 1950 Census of the Americas has proposed for study in the Latin American countries : (1) the utilization of population censuses to test the efficacy of birth registration; (2) the possibility of reorganizing the civil registers before the 1950 census; (3) the calculation of a coefficient of census omissions.

Mr. Loyo remarked on the census tradition of Mexico, but added that as in Europe, the American nations will have to economize in the census operation. He noted that interchange of technical experience will be an important factor in achieving the most efficient operation.

Mr. Mansour suggested that all nations participating in the 1950 censuses should receive guidance from the UN and its Specialized Agencies on problems concerning the accuracy of results. He proposed that there should be a recommendation to national govermennts that legislation be passed to enforce the census taking operations, and establishing secrecy procedures.

Mr. Weiss commented that:

"Any inventory of agricultural resources, be it that of livestock, crop yield, or farming in general, must consider the dependence of all agricultural activities on external forces that act in time and space, that is in a four-dimensional continuum.

"Care must be taken to exclude abnormal events such as inundations, droughts, the occurence of unusual losses due to extraordinary severe hailstorms, plant diseases, or cattle pests, because in all these instances we cannot rely upon the law of great numbers which is only applicable in the case of random distribution of discrete universes.

"The spatial distribution of agricultural activities introduces an element which, though important also in other censuses such as population, manufactures, etc., never reaches such decisive significance as in any instance where land utilization is involved.

"It is a strange fact that statisticians have been more concerned with distribution in time than with the equally important distribution in space. As a result we are in a position to determine time series with some precision while we are unable to deal with space in an equally exact manner....Neither measurements of central tendencies nor measurements of dispersion reveal the one factor which is of so great importance in any appraisal of land utilization, namely the evenness or unevenness of distribution......."

Mr. Lieu pointed out that China will need a great deal of technical assistance in preparation for taking a census of agriculture. He suggested, also, that in China it might be advisable to use sampling methods instead of attempting a complete enumeration.

Mr. Yates supported Mr. Lieu's plea for the full use of sampling methods especially (1) to obtain production estimates; (2) to be more sure of obtaining honest answers; (3) to make more efficient use of skilled enumerators; (4) to achieve the maximum economy of operation.

Mr. Taeuber indicated that the FAO is planning to organize training programs and in other ways to utilize the available technical experience.

Dr. Dedrick discussed training programs which have been set up in the U.S. He pointed out to Mr. Yates that, while he supported the use of sampling, he was aware that a sample survey poorly taken can do irreparable harm to the results, to the agency, and to statistics itself.

Mr. Pelayo also participated in the discussion.

SAMPLING THEORY

Friday, September 12, at 2:00 p.m.[*]

CHAIRMAN :

Will Feller

THE INFLUENCE OF AGRICULTURAL RESEARCH STATISTICS ON THE DEVELOPMENT OF SAMPLING THEORY

by Frank Yates

Rothamsted Experimental Station, England

When I was invited to contribute to the present discussion I was asked if I would be willing to give a general paper on the development of sampling theory, omitting only the recent developments in the United States, which are to be covered by Professor Cochran's paper. Since time was short, however, I decided that it would be better if I dealt mainly with that aspect of the subject with which I am most fully acquainted, namely the contributions of agricultural and biological statistics to sampling theory. The ground to be covered even under these more limited terms of reference is very considerable, and by limiting the scope of the paper I am enabled to give somewhat more detailed treatment of this aspect of the subject.

I understand that it is the task of this session to deal with the development of sampling theory, and that the practice and applications will be dealt with elsewhere. Some reference to practice must however be made, because the technique of sampling has reached its present state by a constant interplay of theory and practice. Indeed I believe that no branch of statistics is likely to develop very far without such interplay. Certainly in all the fields with which I have been associated progress has occurred in this way.

---

[*] Joint meeting with The Econometric Society (Arranged by the American Statistical Association).

The development of sampling theory in biological statistics (and under this head I include agricultural research statistics) has been largely a byproduct of advances in related fields. When confronted with sampling problems workers have tended to use and adapt techniques with which they are already familiar. In particular, the theory of experimental design (and the associated theory of the statistical analysis of the results of experiments) has had a powerful influence on sampling theory which I will now proceed to trace.

THE DESIGN OF EXPERIMENTS AND THE ANALYSIS OF VARIANCE

The principles which form the basis of the modern theory of the design of experiments were developed by R. A. Fisher and his co-workers at Rothamsted in the years immediately following the first world war. Six principles are involved: replication, local control, randomization, splitting of plots for subsidiary treatments, factorial design, and confounding. Only two of these, randomization and confounding, were new. The remainder had all, at one time or another, been used in agricultural field trials, though the recognition of their interrelation is due to the Rothamsted school.

A new technique, that of the analysis of variance, was developed concurrently, in order to provide estimates of the errors to which the results of the experiments were subject. The interrelation between the design and analysis was recognized. Parallel development of the theory of tests of significance, and in particular the *t* and *z* tests, enabled the estimates of error to be used in exact tests of significance.

APPLICATION TO THE SAMPLING OF THE PLOTS OF AGRICULTURAL FIELD TRIALS

Efficient application of the above principles to agricultural field trials requires the use of a large number of small plots in each trial. The full harvesting of these numerous plots created difficulties, and it was natural that the possibilities of sampling the individual plots for yield by harvesting small areas or single plants should be explored.

It was also natural, since the workers concerned were fully imbued with the need for randomization in the design of their experiments, that they should insist on the selection of these small areas or plants at random. This both avoided bias and enabled a valid estimate of the sampling error to be obtained. The principle of local control was also called into play for the reduction of sampling error by the division of each plot into a number of parts (*e.g.*, halves or quarters), and the selection of an equal number of sampling units from each part. This corresponds to the "stratification" used in economic and sociological sampling. The analysis of variance provided a ready machinery by which the sampling error could be calculated in such cases. It led naturally to the elimination of the differences between parts of a plot (which, when there are an equal number of sampling units in each part, do not affect the sampling error), to the pooling of the estimates of error from each plot or part of a plot and to the comparison of the sampling and experimental errors.

Thus, in a variety trial with five replicates of each of four varieties, if 12 sampling units were taken from each plot, 3 from each quarter plot, mechanical extension of the analysis of variance ordinarily employed for the analysis of replicated experiments automatically led to the table:

	Degrees of freedom
Blocks	4
Treatments	3
Experimental error	12
Subtotal	19
Quarter plots	60 (3 D.F. from each plot)
Sampling error (within quarter plots)	160 (2 D.F. from each quarter plot)
Total	239

The first reported investigations into the sampling of agricultural experiments for yield were made by Clapham (1929, 1931) and Wishart and Clapham (1929). The latter paper, dealing with the sampling of potatoes, introduced the important additional principle of complex sampling units, each sampling unit consisting of every 20th plant in each row, the pair of sampling units chosen being determined by random selection of two numbers between 1 and 20. Considerable investigation was undertaken about this time on the relative efficiency of various types of complex sampling unit.

Here again the background of experimental design and the analysis of variance provided the correct solution to the problem of estimating the sampling error of complex units. It was immediately recognized that the variation of the measurements appertaining to the separate members of each unit was irrelevant in the calculation of the sampling error—all that was required was the mean or total of each unit, just as, in the estimation of experimental error, only the total yield of all the plants on each plot is required. Thus the whole complicated structure of correlation was done away with.

The sampling of agricultural field trials was not limited to determination of yield. Sampling observations were frequently taken to determine various characteristics of the growth of the plant, including such things as leaf area, chemical composition, etc., at various stages of maturity. These applications are inherently more interesting than is sampling for yield, since in cases where the observations demand detailed measurements on individual plants, or their destruction, sampling is a necessity, whereas in the case of yield sampling can always be replaced by complete harvesting if this appears desirable.

One important series of observations of this type was that carried out under the Agricultural Meteorological Scheme, in which the growth of two standard varieties of wheat sown in replicated plots at several centers was observed, with a view to determining the effect of meteorological factors (Barnard, 1936). The observations ran for seven years, when they were interrupted by the war. They had an important influence on the development of crop forecasting. Examination of the results accruing during the first few years, for example, suggested that the crop itself might be used as an integrator of the effects of the weather that had occurred up to the date of the observations (Yates, 1936).

This series of observations also gave rise to some interesting investigations into the types of bias that arise from faulty methods of selection in work of this type—investigations which have been of considerable importance in connection with the estimation of the yields of crops from small samples at harvest (Yates and Watson, 1934; Yates, 1935).

EFFICIENCY OF SAMPLING

In the early investigations into the sampling error of agricultural field experiments, no formal study of the efficiency of the sampling relative to the accuracy of the experiment as a whole was made. By 1935, however, a considerable volume of results of actual sampling of cereals experiments had been accumulated and these served as the basis of investigation into this subject (Yates and Zacopanay, 1935). It was recognized that the governing factors are the ratio of the sampling to the experimental error per plot and the ratio of the incremental cost of sampling to other experimental work. Knowing the sampling and experimental errors per plot, the degree of replication required to attain a given accuracy in the final results with any given intensity of sampling per plot can be determined. The minimization of the corresponding cost function serves to determine the optimal amount of sampling under given cost and labor conditions, and also to determine under what circumstances sampling of such experiments is preferable to full harvesting. It is interesting to note that this investigation led to a reversion to full harvesting in circumstances in which sampling had previously been employed, and also to the investigation of an alternative method of sampling, namely sampling for the ratio of grain to total produce, the latter being weighed *in toto* for each plot—an excellent illustration of the importance of investigating the efficiency of sampling processes.

A later and more elaborate investigation of the efficiency of sampling was undertaken by Mahalanobis (1944) in connection with the development of sampling methods for estimating crop acreages and yields in India. Mahalanobis determined a general law of variability of the material sampled, similar to that determined independently by Fairfield Smith (1938) for the variation in fertility of experimental areas in an investigation into the optimal size of plot for agricultural experiments. On the basis of this law of variability and a careful costing of the actual processes of sampling and statistical reduction of the results, Mahalanobis determined a cost function the minimization of which gave him a basis for determining the optimal distribution of his sampling units.

USE OF ANCILLARY INFORMATION CONCERNING THE POPULATION

Sampling for yield by the determination of the grain/total-produce ratio introduced a further problem of considerable theoretical interest, and also of great practical importance in the sampling of economic and sociological material. This is the problem of the estimation of the mean or total of the population, and the corresponding sampling errors, from a ratio determination (here the grain/total-produce ratio) by sampling, in conjunction with some ancillary measurement known for the whole population (here total produce).

If the grain/total-produce ratio is calculated separately for each sample, and the mean of these ratios is taken, the estimation of the sampling error of this mean ratio can be determined in the manner adopted for direct sampling for yield. The estimate of the yield of grain on a plot given by the product of the mean ratio and the yield of total produce may, however, be biased. If, for example, areas with a high yield of total produce also on the average have a high grain/total-produce ratio, an underestimate of the yield will be obtained. This bias can be avoided if the yields of grain and total produce are summed separately for all the sampling units of a plot and the ratio of these totals is used to determine the yield of grain. Symbolically the ratio is estimated from $\Sigma y/\Sigma x$ instead of from $\frac{1}{n}\sum\left(\frac{y}{x}\right)$. The variance of the ratio of the totals can be calculated from the variances and covariance of x and y. This problem was discussed by Cochran (1940), who also gave a second approximation to the variance of such a ratio.

Ancillary information of the above type can be used in another way. The regression of the observed variate y on the ancillary variate x can be determined, and a correction can be applied to the mean value of the observed variate in the sample to allow for the difference between the mean of the ancillary variate in the sample and in the population. Symbolically the corrected estimate is given by

$$Y = \bar{y} - b(\bar{x} - \bar{X})$$

The use of the ratio method is equivalent to taking a value of b equal to $\bar{y}/\bar{x}$ instead of determining from the observations themselves the value of b which will give the most accurate estimate. An application in the British Census of Woodlands to the estimation of standing timber from actual measurements of a sample of stands, and eye estimates of all stands, is described by Cochran (1939).

Here again developments in the theory of regression analysis in the biological field contributed to the correct solution of the problem. In particular, the analysis of covariance, first developed as an extension of the analysis of variance in order to adjust the results of replicated experiments for ancillary information, such as plant number, gave the appropriate method of estimating b in stratified samples.

Another interesting application of regression analysis was developed by Boyd (1947) in connection with the use of eye estimates for crop estimation. (The work was actually carried out in 1940 and 1941.) In order to correct for the inevitable bias in the eye estimates, x, (taken on a random sample of fields), he obtained subsequent information on the harvested yields, y, of a subsample of these fields. This subsample, however, was not random, being conditioned by the availability of the information. In this case an unbiased estimate of the yield is given by

$$Y = \bar{y} - \frac{1}{b'}(\bar{x} - \bar{X})$$

where $\bar{y}$ and $\bar{x}$ refer to the subsample and $\bar{X}$ to the full sample and b' is the regression of x and y.

PURPOSIVE SELECTION AND QUOTA SAMPLING

Purposive selection—that is the selection of the sample in such a manner that certain known quantitative characteristics of the population are reproduced in the sample—and its modern variant, quota sampling, introduce a number of interesting theoretical problems. In the memorandum on sampling prepared by a Commission of the International Statistical Institute (1926) purposive selection figured largely but there was no discussion of the question of what methods of selection might be used to obtain samples having the required values of the known quantitative characteristics without distortion of the other characteristics of the population. In these circumstances it was not surprising that purposive selection led to some bad failures, and the method was severely criticized by Neyman (1934). The Fisherian School, conscious of the simplification introduced into statistical deductions from the results of experiments by the proper use of randomization, also did not favor the method, either in sampling or in the analogous situation in experimental design, in which the plots or animals assigned to the different treatments could be balanced for some known quantitative characteristic such as preliminary yield or initial weight. They pointed out that if the sample was otherwise random the sampling or experimental errors could be estimated by the use of covariance analysis, but that results of equal precision would be obtained from ordinary random samples adjusted by the covariance procedure.

Nevertheless, provided proper methods of selection can be evolved, the use of a balanced sample has an advantage in that no adjustment of the sample means by regression is necessary. This is important in the sampling of economic and sociological material, particularly when large numbers of variates are involved, when mechanical computing methods of the punched card type are employed, and when repeated samplings of the same material have to be made.

In my review before the Royal Statistical Society (1946), I suggested the procedure of random selection with replacement, if the balance was thereby improved, of members of the original sample successively by further randomly selected members. The theoretical points involved were discussed by Anscombe and Quenouille (1946), and Quenouille has subsequently undertaken further work on the problem which will be published shortly.

SAMPLING UNITS AND STRATA OF DIFFERING SIZE AND VARIABILITY

In most of the early sampling problems encountered in agricultural and biological research sampling units of approximately equal size and variability could be employed. In economic and sociological material, however, the natural sampling units often vary widely in size, *e.g.*, households in surveys of human populations, farms in agricultural economic surveys. In such cases, the variability is also not constant, being related to size and other characteristics of the sampling units. The methods appropriate to sampling strata and units of equal size and variability require considerable extension and modification before they can be really useful for economic and sociological material.

This work has made active progress on both sides of the Atlantic during and subsequent to the war (Cochran 1940, 1942; Hansen and

Hurwitz, 1943; Yates, 1946). Hansen and Hurwitz, in particular, have considered the problem of sampling stratified populations in which each stratum contains a small number of primary sampling units of large and varying size. By sampling the primary units within each stratum with probabilities proportional to the sizes of these primary units, and subsampling within the primary units with probabilities so chosen that the over-all probability of selection of a secondary unit is constant for all primary units and all strata, they ensure that the ordinary estimate of the ratio provided by $\Sigma y/\Sigma x$ is both unbiased and free from variability due to variations in size of primary units within strata.[1]

On the agricultural side, graphical methods have been developed, in the course of sampling of agricultural fields for wireworm infestation, in order to provide estimates of sampling error when the variability depends on some property of the sample such as the mean. This is applicable to cases in which the number of degrees of freedom available for the estimation of error is too few for accurate estimation of the sampling error of each sample separately and is analogous to the pooling of estimates of error in the analysis of variance when equal variability can be assumed (Finney, 1941; Yates and Finney, 1942).

USE OF A VARIABLE SAMPLING FRACTION

The possibility of varying the intensity of sampling when differing degrees of accuracy are required on different parts of the material is so self-evident as scarcely to require formalization. The conscious use of such variation of intensity in order to obtain the most accurate over-all values, on the other hand, is relatively recent. The subject was thoroughly discussed by Neyman (1934) and Sukhatme (1935). Neyman pointed out that for maximum precision on the population values, the sampling fractions of the different strata should be chosen proportional to the within-strata standard deviations of the sampling units:

$$\frac{f_1}{\sigma_1} = \frac{f_2}{\sigma_2} = \frac{f_3}{\sigma_3} = \ldots$$

Neyman, however, proposed that the values of the σ, usually unknown, should be determined by preliminary sampling. This made the whole method cumbersome and consequently of limited practical utility.

Fortunately in economic material, in which the use of a variable sampling fraction is of most value, the variability is usually a function of some other characteristic of the material which is known or can be ascertained in advance, and which can be used as a basis of stratification. Very frequently, in particular, the standard deviation is approximately proportional to the size of the unit concerned, and stratification by size, with sampling fractions approximately proportional to the mean size of each size group, leads to very great gains in efficiency. This procedure was used, for example, in the analysis (carried out in 1943-44) of the National Farm Survey of England and Wales (1946). In order to give adequate representation of the smallest size group, and to give somewhat increased precision on statistics of number of farms, the sampling fractions chosen were not quite proportional to mean size. Nevertheless in statistics involving area, which were the main concern of the analysis, there was a gain in precision of three and a half times that which would have been obtained by use of a constant sampling fraction.

---

[1] I regret that in my reply to the discussion on my 1946 paper I failed to recognize the importance and originality of this contribution. I am glad to have this opportunity of correcting my error.

SYSTEMATIC SAMPLING

In the sampling of material which is distributed either temporally or spacially in one or more dimensions, it is frequently simpler to take sampling units at equal intervals (either of time or space), than to locate each sampling unit at random. It was early perceived that in material subject to a more or less continuous type of variation, such sampling (called systematic sampling) was likely to be more accurate than stratified random sampling of the same intensity. The diagrammatic representation of such material, by graphs or maps, is also improved.

By abandonment of the principle of randomization, however, the whole theoretical basis of the estimation of sampling error is destroyed, just as in replicated experiments lack of randomization prevents a valid estimate of experimental error or valid tests of significance.

In experimental work there can be little question that the slight gains in precision and convenience resulting from certain types of systematic design are not sufficient to compensate for the loss of valid estimates of error and other advantages accruing from randomization, and the need for randomization is now almost universally accepted (Yates, 1939). In sampling, however, the issue is not so clear-cut, and there is therefore a case for investigating how far it is possible to estimate the sampling errors in systematic sampling, and what are the appropriate methods. Such investigations have been proceeding in recent years, both in America (see in particular, Cochran, 1946) and Great Britain (Yates, 1946). A full account of my own investigation has not yet been published,* but certain conclusions may be briefly summarized here. Firstly, if the material is such that there is any periodicity, either of the true periodic or autoregressive type, systematic sampling is obviously dangerous and must be avoided. Secondly, though it is relatively easy to determine upper limits to the sampling errors (except in material subject to periodicities) such upper limits are liable to be considerably in excess of the true sampling errors, which can only be determined with certainty by rather extensive supplementary sampling, and, thirdly, that the use of certain simple end corrections to eliminate the effects of trend can be of considerable importance.

---

* Now published (Yates, 1948)

Résumé

Le développement de la théorie des sondages au cours des vingt-cinq dernières années a été étroitement associé à la recherche statistique agronomique. La présente communication expose les fruits de cette association (à l'exclusion des développements récents aux Etats-Unis, traités par le Professeur Cochran) et décrit leur relation avec la théorie des sondages dans les domaines économique et sociologique.

Les contributions principales à la théorie, qui généralement ont résulté de le soluton de problemes pratiques, ont été les suivantes:

1. *Effet de restrictions sur le choix au hasard, les unités de sondage complexes, les sondages d'ordre multiple.*

La théorie du plan d'expériences et l'analyse de la variance ont rendu plus clair le rôle de l'élément aléatoire dans les sondages. Il a été reconnu que l'erreur d'échantillonnage ne pouvait être estimée avec certitude, et que l'absence complète d'erreurs systématiques dans les moyennes empiriques ne pouvait être assurée, que lorsque quelque élément de hasard intervient dans le choix, mais que maints types de restrictions étaient possibles dans ce cadre. L'usage d'unités de sondage complexes et la méthode des sondages d'ordre multiple (sub-sampling), était couvert par la théorie qui, cependant, n'avait été initialement développée que pour traiter des unités de sondage de dimension et de variabilité approximativement égales. L'analyse de la variance fournissait des méthodes appropriées pour l'estimation de l'erreur d'échantillonnage dans de tels cas (Clapham, Wishart et autres).

2. *Efficacité.*

Les lois qui gouvernent l'intensité de l'échantillonnage requise en vue d'une efficacité maximum du travail expérimental furent déterminées en se référant spécialement aux relations entre le coût du travail d'échantillonnage et du travail complémentaire (Yates). L'idée de la

minimisation du coût et de l'usage d'une loi généralisée de variabilité pour le sondage spatial a été développée en rapport avec les essais d'estimation de récolte (Mahalanobis).

3. *Usage de renseignements accessoires concernant la population.*

Dans maints cas des données au sujet de certaines variables aléatoires sont connues exactement (d'après les données de recensements, etc.) pour l'ensemble de la population soumise au sondage, des renseignements sur d'autres variables devant être determinés par échantillonnage. Les valeurs théoriques de ces autres variables peuvent alors être déterminées soit par l'usage de rapports (ou pourcentages) soit à l'aide de la régression (analyse de la covariance). Les méthodes appropriées d'estimation et les erreurs d'échantillonnage qui leur sont associées ont été déterminées (Cochran et autres).

4. *Choix raisonné et échantillonnage par quotas.*

La théorie du contrôle d'un caractère quantitatif en adaptant l'échantillon par rapport à la variable accessoire a été étudiée, et des méthodes de sélection qui échappent aux objections qu'on peut soulever au choix raisonné et à l'échantillonnage par quotas tels qu'ils sont généralement appliqués ont été proposées (Yates et autres).

5. *Unités de sondage de dimensions et variabilité différentes.*

La façon de traiter des unités de sondage de dimensions différentes a été rendue plus claire. Des méthodes graphiques d'estimation de l'erreur ont été développées pour être utilisées dans les cas où la variabilité dépend de quelque autre propriété de l'échantillon, par exemple de la moyenne (Finney).

6. *Usage d'une fraction variable d'échantillonnage.*

Dans plusieurs cas, dans le domaine économique en particulier, l'efficacité de l'échantillonnage est considérablement accrue par l'usage d'une fraction variable d'échantillonnage. La théorie sous-jacente à son usage a été développée, et sa valeur démontrée en pratique (Neyman, Sukhatme, Yates.)

7. *Echantillonnage systématique.*

Les problèmes que posent l'estimation des récoltes et les relevés forestiers ont donné lieu à des recherches étendues dans le domaine de la théorie des sondages systématiques. Une grande partie de ces recherches n'a pas encore été publiée (Cochran, Yates, Quenouille et autres).

References

Anscombe, F. J., and Quenouille, M. H., "Contribution to the Discussion on F. Yates, 'A Review of Recent Statistical Developments in Sampling and Sampling Surveys'," *Journal of the Royal Statistical Society*, Vol. 109 (1946), pp. 38-39.

Barnard, M. M., "An Examination of the Sampling Observations on Wheat of the Crop-Weather Scheme," *Journal of Agricultural Science*, Vol. 26 (1936), pp. 456-487.

Boyd, D. A., "Eye Estimation of Wheat Yields in Hertfordshire" (in preparation).

Clapham, A. R., "The Estimation of Yield in Cereal Crops by Sampling Methods," *Journal of Agricultural Science*, Vol. 19 (1929), pp. 214-235.

Clapham, A. R., "Studies in Sampling Technique: Cereal Experiments," *Journal of Agricultural Science*, Vol. 21 (1931), pp. 366-371.

Cochran, W. G., "The Use of the Analysis of Variance in Enumeration by Sampling," *Journal of the American Statistical Association*, Vol. 34 (1939), pp. 492-510.

Cochran, W. G., "The Estimation of the Yields of Cereal Experiments by Sampling for the Ratio of Grain to Total Produce," *Journal of Agricultural Science*, Vol. 30 (1940), pp. 262-275.

Cochran, W. G., "Sampling Theory When the Sampling Units Are of Unequal Sizes," *Journal of the American Statistical Association*, Vol. 37 (1942), pp. 199-212.

Cochran, W. G., "Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 164-177.

Fairfield Smith, H., "An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops," *Journal of Agricultural Science*, Vol. 28 (1938), pp. 1-23.

Finney, D. J., "Wireworm Populations and Their Effect on Crops, " *Annals of Applied Biology*, Vol. 28 (1941), pp. 282-295.

Hansen, M. H., and Hurwitz, W. N., "On the Theory of Sampling from Finite Populations," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 333-362.

Jensen, A., "Report on the Representative Method in Statistics," *Bulletin of the International Statistical Institute*, Vol. 22 (1926), pp. 359-380.

Mahalanobis, P. C., "On Large-Scale Sample Surveys," *Philosophical Transactions of the Royal Society of London*, series B, Vol. 231 (1944), pp. 329-451.

Ministry of Agriculture, *National Farm Survey of England and Wales: A Summary Report*, 1946. (H. M. S. O.)

Neyman, J., "On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, Vol. 97 (1934), pp. 558-606.

Sukhatme, P. V., "Contribution to the Theory of the Representative Method," *Supplement to the Journal of the Royal Statistical Society*, Vol. 2 (1935), pp. 253-268.

Wishart, J., and Clapham, A. R., "A Study in Sampling Technique: the Effect of Artificial Fertilizers on the Yield of Potatoes," *Journal of Agricultural Science*, Vol. 19 (1929), pp. 600-618.

Yates, F., and Watson, D. J., "Observer's Bias in Sampling Observations on Wheat," *Empire Journal of Experimental Agriculture*, Vol. 2 (1934), pp. 174-177.

Yates, F., and Zacopanay, I., "The Estimation of the Efficiency of Sampling, With Special Reference to Sampling for yield in Cereal Experiments," *Journal of Agricultural Science*, Vol. 25 (1935), pp. 545-577.

Yates, F., "Some Examples of Biased Sampling," *Annals of Eugenics*, Vol. 6 (1935), pp. 202-213.

Yates, F., "Crop Estimation and Forecasting : Indications of the Sampling Observations on Wheat," *Journal of the Ministry of Agriculture*, Vol. 43 (1936), pp. 156-162.

Yates, F., "The Comparative Advantages of Systematic and Randomized Arrangements in the Design of Agricultural and Biological Experiments," *Biometrika*, Vol. 30 (1939), pp. 440-469.

Yates, F., and Finney, D. J., "Statistical Problems in Field Sampling for Wireworms," *Annals of Applied Biology*, Vol. 29 (1942), pp. 156-167.

Yates, F., "A Review of Recent Statistical Developments in Sampling and Sampling Surveys," *Journal of the Royal Statistical Society*, Vol. 109 (1946), pp. 12-43.

Yates, F., "Systematic sampling," *Philosophical Transactions of the Royal Society of London*, series A, Vol. 241 (1948) pp. 345-377.

RECENT DEVELOPMENTS IN SAMPLING THEORY IN THE UNITED STATES

by W. G. Cochran

University of North Carolina (United States)

1. Introduction

In this country the investigation of the theory of sample surveys is of recent origin. The increased use of sampling methods during the late thirties by public opinion polls, market research agencies, and government bureaus took place with relatively little reliance on the services of the mathematical statistician and did not at first give rise to extensions of the underlying theory. Thus the period under review, from about 1939 onwards, contains practically all the development in theory that has occurred here. Discoveries in theory during the period have already profoundly influenced sampling practice and have resulted in an appreciation of the utility of theoretical investigations. These consequences are due largely to the fact that the small group of persons responsible for the developments in theory are also either actively engaged in the taking of sample surveys, or are closely connected with this work.

2- General Trends in Research

The chief guiding principle in current research work is that the object in taking a sample is to obtain estimates of maximum accuracy for a given cost of taking the sample, or alternatively to obtain estimates with a specified degree of accuracy for minimum cost. Since in practice sample estimates can usually be regarded as approximately normally distributed, maximum accuracy is identified with minimum error variance in the sample estimate. The principle is not new: it has been used, for instance, by Yates and Zacopanay (1935) and by Neyman (1938). Nevertheless, its implications were not at first realized either by the practical sampler or by the statistician. The former was inclined to aim at getting the largest size of a sample for a given cost, without appreciating fully that a large sample is itself no guarantee of accuracy. The statistician, on the other hand, tended to base his recommendations on considerations of accuracy alone, perhaps from a reluctance to become involved in cost accountancy. Indeed, knowledge of the relative costs of different sampling methods still lags behind knowledge of their relative accuracies, though a beginning has been made.

If accuracy alone is taken into account, sampling units that are small and widely scattered very frequently appear to be best. A computation of costs may show, however, that a larger sampling unit gives equal accuracy more cheaply, while if subsampling is used, quite large units may be economical. Investigations on optimum size of unit will be reported later (section 6).

The principle has also brought out the importance of careful planning of a sample. In order to apply the principle one must know, before the sample is taken, the costs of the methods of sampling under consideration, the methods of estimation that will subsequently be used, and the sampling variances of the resulting estimates. These facts will not all be known exactly, but serviceable estimates of them can often be made from the results of previous surveys. This approach has led to research on cost functions (sections 5.2, 5.5, 6.2, 7.2), on the properties of different methods of estimation (section 4), and on sampling error formulae for more complex types of sampling (section 5.3,7.1). Attempts have also been made to obtain error formulae for techniques such as systematic sampling (section 8) which though intuitively attractive have not as yet proved amenable to a treatment by standard statistical methods.

It is known that no valid sampling error formula exists unless the selection of the sample is to a certain extent determined by the laws of probability through the use of an objective method of randomization. Such random selection cannot be achieved without something equivalent to a listing of the sampling units in the population. This fact often imposes a serious obstacle to the use of theoretically sound sampling methods, particularly with human populations, where usually no listings are available. The method of *area sampling* represents a major achievement towards overcoming this difficulty. The sampling unit is a compact area of land. A listing of the units in the population is obtained by constructing these areas so that they completely cover a map that shows the population to be sampled. In the Master Sample of Agriculture, designed primarily for farm surveys (King and Jessen, 1945), every county in the United States was divided in this way into areas. The areal sampling units average about 2½ square miles and contain from 4 to 8 farms on the average. Important administrative problems were to devise a rule so that every farm "belongs" to one and only one area, and to provide interviewers with aids (*e.g.*, aerial photographs) to the correct identification of the boundaries of the areas. In large towns the unit is usually a city block, which can also be outlined on a map (for discussion see Hansen and Hauser, 1945). While detailed accounts of the method will not be given, some research involving area sampling is presented in sections 6 (for rural surveys) and 7 (for urban surveys).

In the presentation which follows, I have tried to select the easiest order for the various topics, rather than a chronological order or one of presumed importance. Further, one cannot help being struck by the similarity to developments in India and Britain, which have been reported by Mahalanobis (1944) and Yates (1946). For this reason I have not in general attempted to assign priority of discovery, especially since numerous methods and results were known for some time before publication. In order to maintain a reasonably uniform notation in this paper, I have often, when writing about some result, had to change the notation in which the author originally presented it. This will be an annoyance to the reader who wishes to study the original, but has seemed unavoidable, since sampling literature is noted for the high variance in its notation. A population variance σ^2 has been defined as the sum of squares of deviations divided by $(N-1)$, where N is the number of units in the population. This will change slightly the appearance of formulae in cases where the author has used a divisor N.

3. EXTENSION OF THE GENERAL PRINCIPLE

On further examination, the principle of maximum accuracy for given cost or minimum cost for given accuracy, is not itself completely satisfactory. The principle assumes that in some way either the cost or the accuracy is fixed in advance. Now the specification of the desired degree of accuracy usually involves some arbitrariness. If a coefficient of variation of 1.5 percent is demanded, the sample will not be regarded as useless should the coefficient turn out to be 1.6 percent. The advance specification of a sum of money that must be spent on the sample is also open to criticism; for the accuracy obtained from this expenditure may be substantially more, or substantially less, than is needed for the use that is to be made of the estimates. Two attempts to utilize a more general principle, in which optimum cost and optimum accuracy are determined simultaneously, deserve notice.

In order to apply the principle, one must be able to answer the question: how much is a given degree of accuracy worth?. Any decisions that are based on an estimate from a sample will presumably be more fruitful if the estimate has a low error than if it has a high error. In certain cases we may be able to calculate, in monetary terms, the loss $l(z)$ that will be incurred in a decision through an error of amount z in the estimate. Although the actual value of z is not predictable in advance, sampling theory may enable us to predict the frequency distribution $p(z,n)$ of z, which for a specified method of sampling will depend on the size of sample n. Hence, the *expected* loss for a given size of sample is

$$L(n) = \int l(z)p(z, n)dz.$$

The purpose in taking the sample is to diminish this loss. If $C(n)$ is the cost of a sample of size n, clearly n should be chosen so as to minimize

$$C(n) + L(n)$$

since this is the total cost involved in taking the sample and in making decisions from its results. Choice of n so as to minimize this quantity will determine both the optimum amount of money to be spent on sampling and the optimum accuracy. The idea is presented here only in its simplest form: it may be extended to cover a choice between different sampling methods.

In the application described by Blythe (1945), the selling price of a lot of standing timber is SV, where S is the price per unit volume, and V is the volume of timber in the lot. The number N of logs in the lot is counted, and the average volume per log is estimated from a sample of n logs. If σ is the standard deviation per log for the sampling method used, the standard deviation of the estimate of V will be $N\sigma/\sqrt{n}$ (ignoring finite population correction).

Suppose that this estimate is made and paid for by the seller. The buyer provisionally accepts the estimate of the amount of timber which he has bought. Subsequently, however, he finds out the correct volume purchased, and the seller reimburses him if he has paid for more than was delivered. If he has paid for less than was delivered, the buyer does not mention the fact. In this situation the seller loses whenever he underestimates the volume, but does not gain when he overestimates it. The situation is artificial (my presentation is slightly different from that of Blythe). My purpose is simply to illustrate the application of the principle to a case that does not require complex mathematics.

When he underestimates the volume by an amount z, the seller loses an amount Sz. Thus we may take $l(z)$ as zero when z is negative and as Sz when z is positive, where z is the amount of underestimation. On the assumption of normal distribution of sampling errors, $p(z,n)$ is the normal distribution with mean zero and variance $N^2\sigma^2/n$. Hence

$$L(n) = \frac{\sqrt{n}}{\sqrt{2\pi}\,N\sigma}\int_0^\infty Sze^{-\frac{nz^2}{2N^2\sigma^2}}dz = \frac{SN\sigma}{\sqrt{2\pi n}}$$

If we suppose further that the cost of measuring the volume of a log is c, the cost function $C(n)$ is cn. The quantity to be minimized is therefore

$$cn + \frac{SN\sigma}{\sqrt{2\pi n}}\cdot$$

Differentiation with respect to n leads to the solution

$$n = \left(\frac{SN\sigma}{2c\sqrt{2\pi}}\right)^{\frac{2}{3}}$$

In the example due to Nordin (1944), a manufacturer takes a sample in order to estimate the size of a market which he intends to enter. If the size is known accurately, the amount of fixed equipment and the production per unit period can be adjusted so as to maximize expected profit. Errors in the estimated size of market will result in choices of these two factors that fall short of the optimum, and lead to smaller expected profit. The sample size n should therefore be such that the addition of an $(n+1)$th unit to the sample increases the profit expectation by exactly the cost of the $(n+1)$th unit.

In many cases it will be difficult to apply these ideas because no way can be found to translate the effect of a sampling error into monetary terms. Moreover, an estimate may be used by different persons for quite diverse purposes. Nevertheless, the question of the standard of accuracy needed in sample estimates has received too little attention and this type of research points in a fruitful direction.

4. Estimation of a Population Total

Rather naturally, persons engaged in sampling have favored methods of estimation that can be computed easily and rapidly. The potentialities of complex methods of estimation have been little explored. The gain in accuracy from a superior method of estimation may, however, often be secured fairly cheaply, since only the final computations are affected, and there are likely to be cases, with important estimates, where quite elaborate calculations would be justified if a substantial increase in accuracy resulted. In this section some sampling properties of two methods of estimation, the *ratio* and the *linear regression* methods, are presented. Both have been in use for some time: the circumstances favorable to each are now, however, better understood. For simplicity, the properties are given only for a purely random sample (that is, a sample drawn so that every group of n distinct units in the population has an equal chance of being the sample).

Probably the most common method for estimating the population total Y_p of some item y is to multiply the sample total Y_s by the inverse N/n of the sampling ratio. This estimate may be described as the *mean per s.u.* estimate. Its variance is

$$(1) \qquad V_{N/n} = \frac{N(N-n)\sigma_y^2}{n}$$

In the other two methods an auxiliary variate x, correlated with y, must be obtained for each unit in the sample. In addition, the population total X_p of x must be known. In practice, x is often the value of y on some previous occasion on which a complete census was taken. The aim in both methods is to obtain increased accuracy by taking advantage of the correlation between y and x.

4.1. The ratio estimate. This estimate, which is simple to compute, is

$$\frac{Y_s}{X_s} \cdot X_p$$

This estimate is, in the technical sense, a "best unbiased linear estimate" if two conditions are satisfied: (i) the relation between y and x is a straight line through the origin, and (ii) the variance of y about this line is proportional to x (Cochran, 1942). In exploratory work, a graph plotting the sample values of y against x is therefore useful in considering whether the ratio estimate is likely to be the best available.

Where these conditions are not satisfied, the distribution of the ratio estimate in small samples has not yet been expressed in convenient terms, despite numerous attempts. Unless condition (i) holds, the estimate is biased (Hasel, 1942), though the bias is usually negligible relative to the sampling error. In large samples the distribution tends to normality with a variance which is approximately

$$(2) \qquad V_{ra} = \frac{N(N-n)}{n} \left\{ \sigma_y^2 + R_p^2 \sigma_x^2 - 2 R_p \rho \sigma_y \sigma_x \right\}$$

where $R_p = Y_p / X_p$ is the population ratio of y to x, and ρ is the correlation coefficient between y and x.

On comparison with (1) it is found that the ratio estimate has a smaller variance than the mean per s.u. estimate if

$$\rho > \tfrac{1}{2} \frac{R_p \sigma_x}{\sigma_y} = \tfrac{1}{2} \frac{(c.v.)_x}{(c.v.)_y}$$

where *c.v.* denotes coefficient of variation. Thus the success of the ratio method depends not only on the size of the correlation coefficient, but also on the *c.v.* values. In fact, if the *c.v.* of x is more than twice that of y, the ratio estimate always has a higher variance than the mean per s.u. If x is the value of y on some previous occasion, the two *c.v.*'s may be about equal. In this case the ratio estimate is superior if ρ exceeds $\tfrac{1}{2}$.

4.2. The linear regression estimate. We first compute from the sample the least squares estimate b of the linear regression of y on x. The estimate of Y_p is then taken as

$$(3) \qquad N\{y_s + b(x_p - x_s)\},$$

where y_s, x_s are sample *means* and x_p is the population *mean* of x. Like the ratio estimate, this estimate is slightly biased, though the bias is unimportant. For samples that are reasonably large, the variance may be taken as

$$(4) \qquad V_{\text{l.r.}} = \frac{N(N-n)}{n} \sigma_y^2(1-\rho^2)$$

The error in this formula, which arises from neglect of the sampling variance of b, is in relative terms about $1/n$. The expression has been shown to remain valid in large samples even if the true relation is not linear or if the variance of y changes with x (Cochran, 1942).

From equations (1) and (4) it is clear that in large samples the regression estimate is always more accurate than the mean per s.u. estimate, except when ρ is zero. Comparison of (4) with (2) shows, after a little manipulation, that the regression estimate always has a smaller variance than the ratio estimate unless

$$(5) \qquad \rho = \frac{R_{\text{p}} \sigma_x}{\sigma_y} = \frac{(c.v.)_x}{(c.v.)_y}$$

in which case the two have equal variances.[1] Equation (5) holds whenever the relation between y and x is a straight line through the origin, so that in this event the regression and ratio estimates are equally accurate. The fact that the regression estimate is as good as the ratio estimate even when the latter is a best unbiased linear estimate is interesting.

The regression estimate is more laborious to compute, principally owing to the work in calculating b. However, if there is an appreciable saving in time, an inefficient estimate of b can be used instead of the least squares estimate. If the estimate of b has efficiency E, $(E<1)$, the fractional increase in the variance of the regression estimate of Y_p is about $(1-E)/nE$. With large n, even a highly inefficient estimate of b causes only a trivial increase in the variance. It should, however, be remembered that with the least squares b, one can get an unbiased sample estimate of $\sigma_y^2(1-\rho^2)$ very quickly, whereas with other estimates of b the "short-cut" calculation of the sample residual mean square does not apply.

---

[1] This result was first brought to my attention by Mr. J. R. Goodman.

Another type of estimate has been used by Jessen *et al.* (1947) in a survey taken to estimate the population of Greece. The relation of y to x was found to be a straight line through the origin, but the variance of y about the line was proportional to x^2 rather than to x. The best estimate of Y_p in this case is $\bar{r}_p X_p$, where $\bar{r}_p$ is the arithmetic mean of the sample values of y/x. Some discussion of the properties of weighted regressions has been given by Hasel (1942) and Cochran (1942).

5. STRATIFIED RANDOM SAMPLING

In stratified random sampling the population is divided into a number of groups or strata, the object being to have little within-group variation in the items to be estimated. The total number of sampling units N_h in the hth group is assumed known, and a random sample of n_h is selected from the group. The principles governing stratified sampling have become familiar through the work of Neyman (1934) and the device is almost universally applied. While no major developments have occurred in the period under review, a number of useful results will be given.

5.1. Deep stratification. The number of strata is limited by the size of sample, since there must be at least one sampling unit in each stratum. It may happen that the size of sample must be kept small for administrative reasons, but that many useful criteria for stratification are available. A method of taking advantage of the criteria with an *incomplete* stratification has been mentioned by Frankel and Stock (1942, p. 80).

The sample was taken as part of a monthly survey of unemployment. The population with which we are concerned contains the 447 "urban" counties in the United States. These were divided into three strata according to county population (low, medium, high). Each stratum was further stratified into three groups by geographic location of the counties (NE, SE, and W), and each of the nine cells was still further stratified by the 1937 percentage of unemployment (low, medium, high). Thus there were 27 strata, which might be shown as follows:

Location	County Population		
	Low	Medium	High
NE	l, m, h[1]	l, m, h	l, m, h
SE	l, m, h	l, m, h	l, m, h
W	l, m, h	l, m, h	l, m, h

[1] 1937 % unemployed: low (l), medium (m), high (h).

Suppose that only *nine* strata are chosen, following the latin square plan below:

3 × 3 LATIN SQUARE

	L	M	H
NE	l	m	h
SE	m	h	l
W	h	l	m

For instance, from the three strata which have low county population and NE location, we select only the stratum with low 1937 percent unemployed (l), and so on. Given the sample means in the nine selected cells, we can estimate from the columns of the square, the average difference between counties with low, medium, and high population, from the rows the average difference between counties with NE, SE, and W locations; and similarly from the letters in the square, the average differences caused by the third classification. If these differences are assumed to be additive, the stratum means for all 27 cells can be estimated. By use of a 3×3 Graeco-latin square, a fourth classification could have been added, giving a representation of 81 strata by only nine cells.

Not enough research has yet been conducted on this method to permit a verdict on its utility. The method will obviously be most successful when each classification shows large effects that are independent of those of other classifications. A test of the method in a survey of the city of Wilmington was made by Tepping, Hurwitz, and Deming (1943). They compare four methods of deep stratification with ordinary stratified sampling and with random sampling. The results were, however, inconclusive.

5.2. Optimum allocation with varying costs. If y_{hs} is the sample mean in the *h*th stratum, the "mean per s.u." estimate of the population total is $\Sigma N_h y_{hs}$, summed over the strata. For a given total size of sample, Neyman (1934) has shown that this estimate has minimum variance if the sample sizes n_h are chosen proportional to $N_h \sigma_h$, where σ_h^2 is the variance in the stratum. This well-known principle has often resulted in marked increases in accuracy over proportional sampling, particularly in sampling business enterprises, where the variance may be immensely greater for large than for small firms.

If we apply the general principle of minimum variance for given total *cost*, the Neyman method of allocation is optimum only when the cost per unit is the same in all strata. The best allocation can be worked out for any particular case when the cost function has been obtained in terms of the n_h. A useful result that is easily remembered occurs when the total cost is of the form $\Sigma c_h n_h$. That is, the cost per unit varies from stratum to stratum, but within each stratum costs are proportional to the size of sample.

The variance of the estimate $\Sigma N_h\, y_{hs}$ is

$$(6) \qquad \sum_h \frac{N_h\,(N_h-n_h)}{n_h}\,\sigma_h^2$$

and this is to be minimized subject to the restriction that $\Sigma c_h n_h$ is fixed. Minimum variance will be found to be obtained when n_h is proportional to $N_h\sigma_h/\sqrt{c_h}$. As would be expected, the effect is to diminish the amount of sampling from costly strata.

5.3. Stratified sampling with ratio estimates. There are several ways in which a ratio estimate of a population total Y_p can be made. One is to make a separate ratio estimate of the total of each stratum, and add these totals. If Y_{hs}, X_{hs} are the sample totals in the hth stratum, and X_{hp} is the stratum total for x, this estimate is

$$\sum_h \frac{Y_{hs}}{X_{hs}}\,X_{hp}$$

It is clear that no assumption is made that the true ratio remains fixed from stratum to stratum: the estimate postulates, however, a knowledge of the separate X_{hp}.

Since sampling is independent in the different strata, the variance is found simply by summation of terms as given in formula (2).

$$(7) \qquad V_{sep.} = \sum_h \frac{N_h\,(N_h-n_h)}{n_h}\left\{\sigma_{hy}^2+R_{hp}^2\,\sigma_{hx}^2-2\,R_{hp}\rho_h\sigma_{hy}\sigma_{hx}\right\}$$

where $R_{hp}=Y_{hp}/X_{hp}$ is the true ratio for the stratum.

An alternative estimate, derived from a single *pooled* ratio, has been used by Hansen, Hurwitz and Gurney (1946). This is

$$\frac{\sum \frac{N_h}{n_h}\,Y_{hs}}{\sum \frac{N_h}{n_h}\,X_{hs}}\,.\,X_p$$

Note that only the population total X_p of the x's need be known. The variance of this estimate works out as

$$(8)\quad \underset{\text{pooled}}{V} = \sum_h \frac{N_h(N_h-n_h)}{n_h}\left\{\sigma_{hy}^2+R_p^2\,\sigma_{hx}^2-2\,R_p\,\rho_h\,\sigma_{hy}\,\sigma_{hx}\right\}$$

This differs from (7) only in that the single ratio $R_p=Y_p/X_p$ replaces the R_{hp}. In order to compare (7) with (8), we may use the result (*loc. cit.* p. 179, footnote) that (8) can be written

$$(9)\quad \underset{\text{pooled}}{V} = \underset{\text{sep.}}{V}$$

$$+\sum\frac{N_h(N_h-n_h)}{n_h}\left\{(R_{hp}-R_p)^2\,\sigma_{hx}^2+2(R_{hp}-R_p)(\rho_h\,\sigma_{hy}\,\sigma_{hx}-R_{hp}\,\sigma^2_{hx})\right\}$$

The last term on the right is usually small (it vanishes if within each stratum the relation between y and x is a straight line through the origin). It follows that unless R_{hp} is constant from stratum to stratum, the use of a separate ratio estimate in each stratum is more accurate. The advantage appears to be small, however, unless the variation in the R_{hp} is violent.

For sample estimates of (7) or (8), we substitute the sample estimates of R_{hp} or R_p in the appropriate places. The sample mean squares s_{hy}^2, s_{hx}^2 are substituted for the corresponding variances and the sample mean covariance for the term $\rho_h\sigma_{hy}\sigma_{hx}$. It will be noted that in general the sample mean squares and covariance must be calculated separately for each stratum.

The optimum allocation of the n_h may be different when a ratio estimate is used than when a mean per s. u. is used. In discussing this point, we shall use (7), on the assumption that in practice (7) and (8) will differ little. Now the quantity inside the bracket in (7) is the variance within the hth stratum of the variate $d=(y-R_{hp}x)$. This variance will be called σ_{hd}^2. If (7) is minimized subject to a total cost of the form $\Sigma c_h n_h$, it is found that the n_h must be chosen proportional to $N_h\sigma_{hd}/\sqrt{c_h}$, whereas with a mean per s. u. estimate, n_h is chosen proportional to $N_h\sigma_{hy}/\sqrt{c_h}$.

In the case where the ratio estimate is a best unbiased linear estimate, σ_{hd} will be proportional to $\sqrt{x}$. The n_h would then be made proportional to $N_h\sqrt{x_{hp}}/\sqrt{c_h}$, where x_{hp} is the stratum mean of x. In other cases, *e.g.* that of estimating retail store sales discussed by Hansen, Hurwitz, and Gurney (*loc. cit.*), the variance of d may be more nearly proportional to x^2. This leads to the allocation $n_h \propto X_{hp}/\sqrt{c_h}$, where X_{hp} is the stratum total of x.

5.4. Effects of errors in the strata totals. It frequently happens in practice that for some desirable type of stratification the strata totals N_h are not known exactly, being perhaps derived from a population count that is out of date, or from another sample. Definite statements about the consequences of basing a stratification upon erroneous weights cannot be made without getting down to particular cases. A few conclusions of a general nature can, however, be drawn.

For simplicity, we consider only the estimation of the population mean per unit: finite population corrections will be ignored and the cost per unit is assumed the same in all strata. If the N_h were known, n_h would be chosen equal to $nN_h\sigma_h \,/\, \Sigma N_h\sigma_h$. The sample estimate of the population mean would be $\Sigma N_h y_{hs}/N$, which may be written $\Sigma W_h y_{hs}$. Its variance (which may be obtained from formula (6)) simplifies to

$$\frac{(\Sigma W_h\sigma_h)^2}{n} \tag{10}$$

Instead of the true stratum proportions W_h, we have estimates w_h. The sample estimated mean is $\Sigma w_h\, y_{hs}$. The first point to note is that this estimate is biased. Its mean value in repeated sampling is $\Sigma w_h y_{hp}$ while the true population mean is $\Sigma W_h y_{hp}$. The bias amounts to $\Sigma(w_h - W_h)y_{hp}$. Consequently, the error variance of this estimate contains two components : the variance about its own mean and the square of the bias. If optimum allocation is used (with, of course, the N_h replaced by their estimates) the first component is $(\Sigma w_h\sigma_h)^2/n$. The total variance is

$$\frac{(\Sigma w_h\sigma_h)^2}{n} + \{\Sigma(w_h - W_h)y_{hp}\}^2 \tag{11}$$

A more general form of this expression was given by Stephan (1941). He points out that the first term in (11) will usually be about the same size as (10)—they are exactly the same if the variance is the same in all strata. The loss of accuracy from incorrect weights thus depends mainly on the size of the bias, which in individual cases might either be small or large. Further, for any given set of erroneous weights, the loss varies with the size of sample taken. This is so because the bias component of the total variance is independent of the size of sample. With increasing sample size, a stage is reached where the bias term predominates, and where the stratification would be less accurate than simple random sampling.

The preceding discussion does not help much in considering whether to stratify in a survey where the weights are known to be in error, because the size of the bias term cannot be predicted. Sometimes a standard error can be attached to the estimate of each N_h, from knowledge of the process by which these were estimated. If the estimates of the N_h are independent, and independent of the y_{hs}, the average value of the bias component of the total variance is roughly (Cochran, 1939)

$$\Sigma(y_{hp} - y_p)^2 \, V(N_h)/N^2, \tag{12}$$

where $V(N_h)$ is the variance of our estimate of N_h. This quantity measures the expected increase in variance due to errors in the N_h.

King, McCarty and McPeek (1942) applied this formula in research directed towards the estimation of yield per acre, protein and test weight of wheat in the wheat belt. They discuss the advisability of stratification by districts within each state. The total acreages N_h for each district were themselves estimated by a sample survey, so that some knowledge of the $V(N_h)$ was available.

In sampling human populations, the use of erroneous strata totals can often be avoided by area sampling, as discussed in sections 7.2 and 7.3.

5.5. The problem of non-response. In many types of survey, there are certain units in the sample from which the desired information cannot be obtained at the first attempt. With human populations, this group may be persons who are not at home, or who do not reply to a mail questionnaire. Similarly, in crop surveys certain fields in the sample may not be ripe when the sampler reaches them. This "non-response" group constitutes an important practical problem. To obtain information from it may require several attempts and be costly. To ignore it may result in a sample that has a bias of unknown dimensions. An ingenious application to this problem of the idea of stratified sampling has been made by Hansen and Hurwitz (1946).

The population is envisaged as containing two strata. One, of size N_1, contains units that provide the information at the first try. The second, of size N_2, is the non-response stratum. The basic idea is that the second stratum should be sampled at a lower rate than the first, since the cost per unit is higher in that stratum. There is, however, the complication that neither the values of N_1 and N_2, nor even the units that fall in the two strata, is known in advance.

The first step, in the simplest case, is to take a random sample of n units. Of these let n_1 be the number that provide the data sought, and n_2 the number in the non-response group. By repeated efforts, the data are later obtained from a random sample of r_2 out of the n_2. If $n_2 = kr_2$, the quantity k is the ratio of the sampling rate in the first stratum to that in the second. The values of n (initial size of sample) and k are chosen so as to give a specified accuracy for the lowest cost.

The cost of taking this sample is

$$c_0 n + c_1 n_1 + c_2 r_2, \tag{13}$$

where the c's are costs per unit: c_0 is the cost of making the first attempt, while c_1 and c_2 are the costs of getting and processing the data in the two strata respectively. Since the values of n_1 and r_2 will not be known until the first attempt is made, the *expected* cost must be used in planning the sample. The expected values of n_1 and r_2 are respectively $W_1 n$ and $W_2 n/k$, where as previously $W_1 = N_1/N$. Thus expected cost is

$$c_0 n + c_1 W_1 n + c_2 W_2 n/k. \tag{14}$$

The estimated population total is taken as

$$\frac{N}{n}\left\{ n_1 y_{1s} + n_2 y_{2s} \right\}$$

and its variance is shown to be

$$\frac{N(N-n)}{n}\sigma^2 + \frac{(k-1)N\,N_2}{n}\sigma_2{}^2 \tag{15}$$

where σ^2 is the variance in the whole population and $\sigma_2{}^2$ is that within the non-response stratum. The first component is the variance that would be obtained if all n_2 in the non-response group were sampled; the second is the increase from sampling only r_2 of the n_2. The quantities n and k are then chosen to minimize (14) for a pre-assigned value of (15).

The solutions depend on the unknowns N_1 and N_2. If fairly close estimates of these can be made from earlier experience, the estimates may be used in place of the unknowns. Even if nothing is known in advance about N_1 and N_2, the authors develop an alternative method that gives in most cases a solution close to the optimum. Extensions to stratified sampling and to ratio estimation are also presented.

6. OPTIMUM SIZE OF SAMPLING UNIT: CLUSTER SAMPLING

In certain types of survey the size of unit may be varied over quite a wide range. In accordance with the general principle, the best size must depend on a balance between the accuracy obtained from a given size, and the cost of sampling with that size. An investigation by Jessen (1942) presents a good illustration.

The population consisted of the farms in Iowa. The unit was a compact area of land delineated on the map, or more accurately all farms defined as belonging to this area. While certain sizes of area (*e.g.*, multiples or submultiples of a square mile) have particular practical conveniences in Iowa, the size will be regarded as capable of continuous variation. For simplicity in presentation, all areas of a given size are assumed to contain the same number of farms, and the variance of any item between farms within areas of the same size is assumed constant over the state. Neither assumption seriously vitiates the conclusions in this case.

6.1. The variance function. For some common farm economic items, Jessen found that the variance W between farms within an area which contains M farms is given by the equation

$$W = AM^{\kappa}, \qquad g > 0, \tag{16}$$

where W is measured on a single-farm basis, and A and g are constants independent of M. In this model W increases as the size of area increases, the curve being concave upwards. This increase is familiar in farm sampling, and may be attributed to the fact that many forces, including weather and soil type, tend to exert a similar influence on farms that are close together.

Suppose that a random sample of n of these areas is selected. The variance that we are interested in is that of the estimated population total or mean: Jessen considers the latter. If B is the variance between areas (on a single-farm basis) the variance of the sample mean per farm is B/Mn, since the sample contains Mn farms (finite population correction is neglected). Thus we require knowledge of B as a function of M.

The relation between B and W is supplied by the following analysis of variance for *all* farms in the state. The number of farms in the state is N, so that the number of areas is N/M.

	d.f.	Mean square
Between areas with M farms	$\left(\frac{N}{M} - 1\right)$	B
Within areas between farms	$\frac{N}{M}(M-1)$	W
Total between farms in the state	$(N-1)$	T

From the additive property of sums of squares, it follows that

$$\frac{(N-M)B}{M} = (N-1)T - \frac{N(M-1)W}{M} \tag{17}$$

It will be noted that T does not depend on M. Substituting for W from (16) and dividing by Mn, we express the variance of the estimated population mean per farm as shown below.

$$V(M,\, n) = \{\, M(N-1)T - N(M-1)AM^{g} \,\}/Mn(N-M) \tag{18}$$

Assuming N large, Jessen simplifies this to

$$V'(M,\, n) = \frac{1}{n}\left[T - (M-1)AM^{g-1}\right] \tag{19}$$

To digress for a moment, the whole population might be regarded as a single sampling unit of size N. Consequently, if formula (16) still holds for such large units, we may substitute $T = AN^{g}$ in (18). This reduces the number of unknown constants to be determined. Hendricks (1944) found this relation to hold satisfactorily for corn and wheat acreage data, and McVay (1947) has applied it to the estimation of the proportion of orchards that grow peaches. It may happen, however, that the formula for W, while remaining valid for the range of values of M that we are likely to consider, does not hold out to the value N. In this case the more general relation (18) should be used. The constants T, A, and g must of course be estimated from data.

6.2. The cost function. Jessen distinguishes two elements of cost. The element c_1Mn comprises costs that vary directly with the total number of farms : thus c_1 contains the cost of an interview and the cost of travel from farm to farm within an area.

The second element, $c_2\sqrt{n}$, measures the cost of travel between the areas. By tests on a map it was found that this cost, for a fixed population, varied roughly with the square root of the number of areas. Total cost is therefore of the form

$$C(M,\, n) = c_1Mn + c_2\sqrt{n}. \tag{20}$$

Equation (19) for the variance is then minimized with respect to variation in M and n, subject to the restriction (20). While there is no neat explicit solution, the numerical solution in a specific case is not difficult.

One form of the theoretical solution for M, obtained after some manipulation, leads to some general conclusions that are of interest.

$$\frac{M}{V'}\frac{\partial V'}{\partial M} = \left\{1 + \frac{4Cc_1M}{c_2^{2}}\right\}^{-\frac{1}{2}} - 1 \tag{21}$$

The left hand side is purely a function of M and the constants T, A, and g: it does not depend on n or the cost factors. Consequently, so far as cost factors are concerned, the optimum M depends only on the quantity $q = Cc_1/c_2^2$. Further, it may be shown that as q increases, the optimum M decreases. But c_1 increases as the length of interview increases, while c_2 decreases if travel becomes cheaper, or if farms become more dense. These facts provide the general conclusion that, other things being equal, the optimum size of sampling unit becomes smaller if (i) the length of interview increases, (ii) travel becomes cheaper, (iii) the elements (farms) become more dense, or (iv) total amount of money available (C) increases.

With urban sampling, the sampling units that suggest themselves are the individual, the household, and the group of households. Hansen and Hurwitz (1942) used the name *cluster* for a group of elements (*e.g.* individuals) that constitute a sampling unit. They present examples which show that for certain items in urban sampling the variance function is quite different from that used by Jessen. In estimating the sex ratio, for instance, the variance B between households is only about half as large as the variance T between individuals, whereas in formula (17) B is always envisaged as being greater than T. The common presence of both husband and wife in the same house introduces a negative correlation between the sexes of members of a household. For this reason, the authors have preferred to express their mathematical model in terms of the intra-cluster correlation coefficient, which may take either positive or negative values.

7. SUBSAMPLING

With large sampling units, a useful device is to sample only a part of each unit that is selected. Thus if the large unit contains M *sub-units*, we might obtain information from m of these, chosen at random. When sub-units within the same unit give closely similar results, common sense suggests that it will not be profitable to spend time measuring a large number of them, at the expense of decreasing the number of units that can be sampled. For instance, in estimating wheat yield and quality by sampling at harvest time, King and McCarty (1941) found that it was advisable to cut only a very small quantity of wheat from each field that was sampled.

The sampling theory is simple when each unit contains the same number M of sub-units, of which m are chosen, and when the variance σ_w^2 between sub-units in the same unit is constant. The variance of the sample mean per sub-unit is

$$\frac{(N-n)}{N}\,\frac{\sigma_b^2}{n} + \frac{(M-m)}{M}\,\frac{\sigma_w^2}{mn}$$

where σ_b^2 is the variance between the *true* means of the units and n units are sampled out of N. An unbiased estimate of this variance is obtained from an analysis of variance of the sample. If B is the mean square between units and W that between sub-units within units, both being on a sub-unit basis, it may be shown that

$$E(B) = m\sigma_b^2 + \frac{(M-m)}{M}\sigma_w^2, \quad E(W) = \sigma_w^2.$$

Consequently, the unbiased sample estimate is

$$\frac{1}{mn}\left[\frac{(N-n)}{N}B + \frac{(M-m)}{M}\frac{n}{N}W\right]$$

7.1. Subsampling when sampling units vary in size. This problem has arisen frequently in recent years especially with human populations, where the unit may be a city block, or a county, and the sub-unit a household or a farm. At first sight a confusing variety of methods of sampling and of estimation present themselves. The appraisal of these methods and the development of new methods is an important contribution by Hansen and Hurwitz (1943).

The methods and their properties can be illustrated for the simple population represented below.

Unit no. (i)	M_i	Sub-unit values	Unit total
1	6	1,1,1,1,1,1	6
2	2	0,0	0
	8		6

The population has only two units. The first has six sub-units, all of which give the value 1 for the item measured. The second has two units, each giving 0. Thus σ_w^2 is zero. The population total (the quantity to be estimated) is 6. It is proposed to select one unit, so that $n = 1$, $N = 2$. Two sub-units will be sampled from the unit chosen. The sample therefore comprises two of the 8 sub-units in the population.

The methods and the results for this population are summarized in Table 1.

TABLE 1

Subsampling Methods With Units of Unequal Sizes.

Method	p_i	m_i	Estimate	Values	Bias*	σ^2
Ia	equal	m	$\frac{N}{n} \cdot \frac{M_i}{m} \cdot Y$	12, 0	U	36
Ib	equal	m	$\frac{M}{m} \cdot Y$	8, 0	B	20
IIa	equal	fM_i	$\frac{N}{n} \cdot \frac{M_i}{m_i} \cdot Y$	12, 0	U	36
IIb	equal	fM_i	$\frac{M}{m_i} \cdot Y$	8, 0	B	20
IIIa	$\frac{M_i}{M}$	m	$\frac{M}{m} \cdot Y$	$8(\frac{3}{4}), 0(\frac{1}{4})$	U	12
IIIb	P_i	$\frac{tM_i}{P_i}$	$\frac{1}{t} \cdot Y$	$9(\frac{2}{3}), 0(\frac{1}{3})$	U	18

* U = unbiased, B = biased.

Method Ia. Choose a unit at random and $m = 2$ sub-units from it. Multiply the sample total Y by M_i/m to give an unbiased estimate of the unit total, and then by N/n, or 2, to give an unbiased estimate of the population total. This estimate is 12 if the first unit is chosen and 0 if the second unit is chosen. The sampling variance of the estimate is 36.

On reflection, the method of estimation leaves something to be desired. The estimate is N(estimated unit total)$/n$. But the variation in the unit *totals* depends on the variation in the sizes M_i as well as on that in the sub-unit values. Even if the latter all had the value 1, this method would give a large variance if the M_i differed greatly. A technique that avoids this type of inflation of the variance is desired.

Method Ib. Same method of sampling. The estimate is the sample mean, multiplied by the number of sub-units in the population: *i.e.*,

MY/m, or in this case $4Y$. The two possible estimates are 8 and 0. Since their mean, 4, does not equal the true population total, the estimate is biased. The error variance (about the true population total) is $\frac{1}{2}(2^2 + 6^2)$, or 20. Despite the bias, this estimate is more accurate than Ia. The reason is that since the estimate is based on the sample *mean* its error is not inflated by variations in the M_i.

Method IIa. Select a unit at random as before, but take $m_i = \frac{1}{2}M_i$ (*proportional subsampling*). This procedure is derived from the idea that large units should in some way receive more representation in the sample than small units. The sample will be of size 3 if the first unit is chosen, and of size 1 if the second is chosen. While it may seem undesirable that the sample size should depend on the particular units selected, over a number of strata the actual size would be near the expected size. The estimate is analogous to that in Ia ; *i.e.*, $\frac{N}{n} \cdot \frac{M_i}{m_i} \cdot Y$. The reader will find that this method gives exactly the same estimates as Ia and hence has the same variance, 36.

Method IIb. Sampling method as in IIa: estimate as in Ib. (MY/m_i). Results are the same as in Ib, the estimate being biased, with variance 20.

The equivalence of IIa with Ia and IIb with Ib always holds when there is no variation within units. When sub-units within the same unit differ, as will happen in practice, Hansen and Hurwitz (1943, p. 357) have shown that Ib will usually be more accurate than IIb.

Method IIIa. The authors propose to give greater representation to large units by drawing the units, not with equal probabilities, as in all previous methods, but with probability *proportional to their sizes.* Thus $p_i = M_i/M$. A probability $\frac{3}{4}$ is assigned to the large unit and $\frac{1}{4}$ to the small unit. As in methods I, two sub-units are sampled from the unit chosen. The estimate is MY/m, or $4Y$. In repeated sampling we get an estimate 8 in $\frac{3}{4}$ of the samples, and one of 0 in the remaining $\frac{1}{4}$. The estimate is unbiased. Its variance is

$$\frac{3}{4}(2)^2 + \frac{1}{4}(6)^2 = 12$$

This value is considerably smaller than that given by any previous method.

By examination of the sampling error formulae for the different techniques, the authors show that method IIIa will give the smallest between-unit contribution to the variance for many types of population that appear to be met in practice. So far as the within-unit contribution is concerned, this is not likely to differ greatly in methods IIIa and

Ib, which is usually the most accurate of the other methods described thus far (method IIIb will be discussed in section 7.2). The between-unit contribution is quite commonly, though not always, much larger than the within-unit contribution. This was the reason for neglecting the latter in the example.

7.2. Case where the sizes are not known. It may happen that the sizes M_i of the units (*e.g.*, number of households in a block or county) are not known when the sample is drawn, though estimates of them may be obtainable from previous data. Insofar as these sizes are needed for sampling or estimation, time and money must be spent in obtaining them. It will be seen from Table 1 that with methods Ia and IIa we require the M_i only *for those units that come into the sample.* With the more accurate methods Ib, IIb, and IIIa, we must count the total M of *all* the units. If counting is expensive, the cheaper cost of methods Ia and IIa may go far to counterbalance their relatively large variances.

An interesting application of an extension of method IIa is presented by the sampling staff of the Bureau of the Census in *A Chapter in Population Sampling* (1947). The primary object of the sample was to estimate the total population in certain congested areas, each containing a city and the surrounding rural area. In the cities the unit was a block and the sub-unit a dwelling-place. The blocks were stratified by estimated size (number of dwelling-places) according to data obtained four years previously. This stratification would remove a considerable part of the variation of the M_i. The sampling ratio n/N and the sub-sampling ratio m_i/M_i were determined so as to minimise for a specified accuracy, the total cost of counting the M_i in the sampled units, enumerating the m_i and tabulating the results. Special methods were devised for trailer camps, institutions, etc., for the rural areas, and for blocks reported as having no dwelling-places in the previous data.

There is another method of sampling (IIIb) which requires a knowlege only of the M_i that are being sub-sampled. It is in a sense a generalization both of methods IIIa and IIa, and is due to Hansen and Hurwitz (1943). It is an application of IIIa in the case where we have only estimates P_i of the relative sizes, though mathematically the P_i can be any set of positive numbers adding to unity over each stratum . We first select a *sampling rate t*, which may vary from stratum to stratum. In the example t is chosen as 2/8, since a sample of size 2 is desired and M is 8. A sampling unit is then chosen with a probability P_i attached to the ith unit. The M_i is counted for the unit chosen and m_i is taken as tM_i/P_i. The estimate is Y/t and therefore does not require a knowledge of M.

In the example, suppose that we estimated the P_i as $\frac{2}{3}$ for the large unit and $\frac{1}{3}$ for the small. The sample sizes work out at 9/4 for the large unit and 3/2 for the small. We shall assume that samples of these non-integral sizes could actually be taken (they could be approximated closely if the M_i were large). The estimates are 9, with probability $\frac{2}{3}$, and 0 with probability $\frac{1}{3}$. The estimate is unbiased (this is always true). The sampling variance is $\{\frac{2}{3}(3)^2 + \frac{1}{3}(6)^2\}$, or 18.

The reader may verify that if $P_i = M_i/M$, the method reduces to IIIa, while if $P_i = 1/N$, the method reduces to IIa. When the P_i differ from M_i/M, the sampling error is increased to some extent through the variations in the M_i, though not greatly if the estimates of size are good.

7.3. Further comments. Methods IIIa and IIIb have been extended to stratified sampling and to ratio estimation (*loc. cit.*). When only one unit is sampled per stratum, the actual drawing of the unit is simple. The P_i, or a convenient multiple of them, are cumulated. If their grand total is G, a random number r is chosen between 1 and G. The unit which contains this number in the list of cumulative totals is the unit selected. Since mathematical complications enter when more than one unit per stratum is to be drawn, it is customary to stratify to the point where one unit per stratum is taken.

With methods Ia or IIa it is highly advisable to stratify by estimated size if the M_i are likely to vary considerably. With methods IIIa and IIIb, however, stratification can be by other factors, since size effects are removed by selection of units with probability proportional to size. It is worth emphasis that with methods Ia, IIa, or IIIb the stratum total numbers of sub-units need not be known exactly. Thus these methods allow stratification without risking the bias that may be introduced by the use of erroneous strata totals, as discussed in section 5.4.

When only one unit is sampled from a stratum, the subsample from that unit should represent the *stratum* rather than the unit. In some cases it is possible to stratify the sub-units in the stratum by a criterion different from that used in forming the original strata. If so, an increase in accuracy may be secured by drawing the subsample so that it is stratified according to this criterion. The sampling error formula for this technique, which is known as *area substratification*, will be found in the same reference.

8. SYSTEMATIC SAMPLING

If the units in a population are numbered from 1 to $N = nk$, a common method of drawing a sample of n is to take a random number between 1 and k, say r, and draw the rth unit and every kth unit thereafter. This type of sample is called a systematic sample. It can be drawn more quickly than a random sample and intuitively seems likely to be more accurate, since it spreads the sample evenly over the population. In some respects it resembles a stratified random sample in which each stratum has k units, and one unit is sampled per stratum. The difference is that in the systematic sample the unit sampled always occurs at the same place in the stratum. Frequently it is used mainly for convenience, the order in which the units are numbered having no particular significance. In this case it may be predicted that the systematic sample will give about the same accuracy as a random sample of the same size. It has, however, not proved easy to make general statements about the accuracy of systematic samples.

The sampling theory was investigated by W. G. and L. H. Madow (1944). They give several formulas for the variance of the sample mean, of which one may be quoted. In our notation, this is

$$\frac{\sigma^2}{n}\left\{\frac{N-1}{N}+\frac{2}{n}\sum_{d=1}^{n-1}(n-d)\,\rho'_{kd}\right\}$$

where σ^2 is the population variance and ρ'_{kd} is the noncircular serial correlation coefficient with lag *kd*, defined by

$$k(n-d)\sigma^2\rho'_{kd}=\sum_{i=1}^{k(n-d)}(y_i-y_p)(y_{i+kd}-y_p).$$

The difficulty is that the accuracy depends on the nature of the lag correlations, such as are caused by any continuous variation in the items y_i.

If the major variation is a linear trend, the Madows show that the systematic sample is more accurate than a random sample, but less accurate than a stratified random sample with one unit per stratum. If the major variation is periodic, the systematic sample is highly accurate if k is an odd multiple of half the period, but is inaccurate if k is a multiple of the period. If the major variation is such that ρ'_i is a positive, monotone-decreasing, and concave upwards function of i, the systematic sample is always at least as accurate, on the average, as the stratified random sample (Cochran, 1946). This type of population was studied because in it the variance within a group of i consecutive units increases as i increases, as often seems to be the case in practice (*e.g.*, Jessen's variance formula given earlier has this property).

Investigations of this type must be supplemented by examination of the types of continuous variation that are encountered in practice. There is reason to believe that frequently the systematic sample will be substantially more accurate than either the random or stratified random samples. For instance, Johnson (1943) studied 13 natural populations in which the observations were the numbers of seedlings in successive feet in a forest nursery bed. Five species were represented. In seven beds containing seedbed stock with high variability, the variance of the systematic sample was about half that of the stratified random sample. In the remaining six beds, which had rather homogeneous transplant stock, the systematic sample was about equal in accuracy to the stratified random sample and considerably more accurate than the random sample. Osborne (1942) found the systematic sample twice to four times as accurate as the stratified sample in estimating the areas under different types of cover (*e.g.*, grass, woodland) from a map. In both investigations the stratified sample had a stratum of size $2k$ with two samples per stratum so as to permit estimation of the sampling error. It appears, however, that the results would be little changed if the stratum size were reduced to k.

For a given population, the variance of the systematic sample mean is likely to be a much less stable function of n than with the random methods of sampling. This is illustrated by a silver maple bed from Johnson's data (L.H.Madow, 1946). Variances of sample means for samples from $n = 10$ to $n = 84$ are shown below : the value of N was 420.

TABLE 2

Variances of Sample Means

(*from L.H.Madow*, 1946)

n	Type of sample Random	 Stratified random*	 Systematic
10	10·29	7·21	4·21
21	4·77	3·00	3·06
28	3·52	2·09	2·42
30	3·26	1·90	0·69
42	2·26	1·29	1·74
60	1·51	0·82	0·26
84	1·00	0·51	1·22

* One unit per stratum.

The rather erratic behavior of the systematic sample is evident. For some sample sizes ($n = 10$, 30, and 60) there are striking gains over the stratified sample; for others ($n = 21$, 28, and 42) the variance is slightly higher than for stratified sampling; while for $n = 84$ the variance is actually higher than with simple random sampling. This type of effect is to be expected. The systematic samples of size 60, for instance, are each composed of two of the samples of size 30. Thus there is a positive correlation between the variances for $n = 30$ and $n = 60$.

It is well known that from the results of a single systematic sample one cannot obtain an unbiased estimate of the sampling variance of the mean that will be valid for any type of population. It should be feasible to develop estimates that will be adequate for practical purposes, given some knowledge of the type of population that is being sampled. With his data, Osborne (1942) obtained good results from a formula $\delta^2 (1-\bar{r}^2)/2n$, where δ^2 is the mean square successive difference $\sum_{i=1}^{n-1} (y_i - y_{i+k})^2/(n-1)$, and $\bar{r}^2$ is the average of the squared correlations between a unit in the sample and all other units in the same group of k units. Further study of actual populations will be profitable.

References

A Chapter in Population Sampling, by the Sampling Staff, Bureau of the Census, 1947.

Blythe, R. H., "The Economics of Sample Size Applied to the Scaling of Saw Logs," *Biometrics Bulletin*, Vol. 1 (1945), pp. 67-70.

Cochran, W. G., "The Use of the Analysis of Variance in Enumeration by Sampling," *Journal of the American Statistical Association*, Vol. 34 (1939), pp. 492-510.

Cochran, W. G., "Sampling Theory When the Sampling Units are of Unequal Sizes," *Journal of the American Statistical Association*, Vol. 37 (1942), pp. 199-212.

Cochran, W. G., "Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 164-177.

Frankel, L. R. and Stock, J. S., "On the Sample Survey of Unemployment," *Journal of the American Statistical Association*, Vol. 37 (1942), pp. 77-80.

Hansen, M. H. and Hauser, P. M., "Area Sampling—Some Principles of Sample Design," *Public Opinion Quarterly*, Vol. 9 (1945), pp. 183-193.

Hansen, M. H. and Hurwitz, W. N., "Relative Efficiencies of Various Sampling Units in Population Inquiries," *Journal of the American Statistical Association* Vol. 37 (1942), pp. 89-94.

Hansen, M. H. and Hurwitz, W. N., "On the Theory of Sampling from Finite Populations," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 333-362.

Hansen, M. H. and Hurwitz, W. N., "The Problem of Non-Response in Sample Surveys," *Journal of the American Statistical Association*, Vol. 41 (1946), pp. 517-529.

Hansen, M. H., Hurwitz, W. N., and Gurney, M., "Problems and Methods of a Sample Survey of Business," *Journal of the American Statistical Association*, Vol. 41 (1946), pp. 173-189.

Hasel, A. A., "Estimation of Volume in Timber Stands by Strip Sampling," *Annals of Mathematical Statistics*, Vol. 13 (1942), pp. 179-206.

Hendricks, W. A., "The Relative Efficiencies of Groups of Farms as Sampling Units," *Journal of the American Statistical Association*, Vol. 39 (1944), pp. 367-376.

Jessen, R. J., "Statistical Investigation of a Sample Survey for Obtaining Farm Facts." *Iowa State College Research Bulletin* 304, 1942.

Jessen, R. J., Blythe, R. H., Kempthorne, O., and Deming, W. E., "On a Population Sample for Greece," *Journal of the American Statistical Association*, Vol. 42 (1947), pp. 357-383.

Johnson, F. A., "A Statistical Study of Sampling Methods for free Nursery Inventories," *Journal of Forestry*, Vol. 41 (1943), pp. 674-679.

King, A. J. and Jessen, R. J., "The Master Sample of Agriculture," *Journal of the American Statistical Association*, Vol. 40 (1945), pp. 38-56.

King, A. J., and McCarty, D. E., "Application of Sampling to Agricultural Statistics with Emphasis on Stratified Samples," *Journal of Marketing*, Vol. 5 (1941) pp. 462-474.

King, A. J., McCarty, D. E., and McPeek, M., "An Objective Method of Sampling Wheat Fields to Estimate Production and Quality of Wheat," *U. S. Department of Agriculture Technical Bulletin* 814, 1942.

Madow, L. H., "Systematic Sampling and its Relation to Other Sampling Designs," *Journal of the American Statistical Association*, Vol. 41 (1946), pp. 204-217.

Madow, W. G. and Madow, L. H., "On the Theory of Systematic Sampling I," *Annals of Mathematical Statistics*, Vol. 15 (1944), pp. 1-24.

Mahalanobis, P. C., "On Large-Scale Sample Surveys," *Philosophical Transactions of the Royal Society of London*, Vol. 231 (1944), pp. 329-451.

McVay, F. E., "Sampling Methods Applied to Estimating Numbers of Commercial Orchards in a Commercial Peach Area," *Journal of the American Statistical Association*, Vol. 42 (1947), pp, 533-540.

Neyman, J., "On Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, Vol 97 (1934), pp. 558-606.

Neyman, J., "Contribution to the Theory of Sampling Human Population," *Journal of the American Statistical Association*, Vol. 33 (1938), pp. 101-116.

Nordin, J. A. "Determining Sample Size," *Journal of the American Statistical Association*, Vol. 39, (1944), pp. 497-506.

Osborne, J. G., "Sampling Errors of Systematic and Random Surveys of Cover-Type Areas," *Journal of the American Statistical Association*, Vol, 37 (1942), pp. 256-264.

Stephan, F. F., "Stratification in Representative Sampling," *Journal of Marketing*, Vol. 6 (1941), pp. 38-46.

Tepping, B. J., Hurwitz, W. N., and Deming, W. E., "On the Efficiency of Deep Stratification in Block Sampling," *Journal of the American Statistical Association*, Vol. 38 (1943), pp. 93-100.

Yates, F., "A Review of Recent Statistical Developments in Sampling and Sampling Surveys," *Journal of the Royal Statistical Society*, Vol. 109 (1946), pp. 12-43.

Yates, F. and Zacopanay, I., "The Estimation of the Efficiency of Sampling with Special Reference to Sampling for Yield in Cereal Experiments," *Journal of Agricultural Science*, Vol. 25 (1935), pp. 545-577.

Résumé

Lorsqu'on effectue un sondage, on vise soit à obtenir le maximum de précision dans les estimations pour une dépense donnée, soit à minimiser le coût d'estimations d'un degré déterminé de précision. Les recherches aux Etats-Unis sont régies par ces principes. Un principe encore plus général a été introduit par *Blythe* et *Nordin.* Suivant ce principe, le coût optimum et la précision optimum sont déterminés simultanément en minimisant le coût de l'enquête et le coût de décisions erronées dûes à l'imperfection des estimations.

Dans la communication sont examinés: 1) les estimations du total de la population, 2) la méthode du sondage aléatoire stratifié, dont quelques nouvelles applications ont été dévouvertes, 3) le choix optimum de l'unité d'échantillonnage, 4) la méthode de sub-échantillonnage, 5) la méthode du sondage systématique, très souvent utilisée en pratique,

SEQUENTIAL ANALYSIS

by Abraham Wald
Columbia University (United States)

Sequential analysis has been developed in the last four years with special reference to the problem of testing statistical hypotheses.

The notion of a statistical hypothesis may be defined as follows: Let X be a random variable whose probability distribution is known except for the values of some parameters involved in the distribution function of X. For example, we may know that the distribution of X is normal but we do not know the mean value or the standard deviation of the distribution. Suppose that k unknown parameters $\theta_1, ..., \theta_k$ are involved in the distribution of X. A statement about the values of the unknown parameters is called a simple hypothesis if it determines uniquely the values of all k parameters. It is called a composite hypothesis if it is consistent with more than one value of some parameter. For example, if there are two unknown parameters θ_1 and θ_2, the hypothesis that $\theta_1 = 2$ and $\theta_2 = 3$ is a simple one, while the hypothesis that $\theta_1 = \theta_2$ is a composite one.

An essential feature of the sequential method for testing a statistical hypothesis H is that the number of observations required by the test procedure is not determined in advance but is dependent on the results of the observations. The nature of the sequential test procedure may be described as follows: A rule is given for making one of the following 3 decisions at each stage of the experiment (at the mth trial for each integral value m):

1) to accept H;
2) to reject H;
3) to continue the experiment by making an additional observation.

Thus, such a test procedure is carried out sequentially. On the basis of the first observation one of the above mentioned 3 decisions is made. If the first or the second decision is made, the process is terminated. If the third decision is made, a second trial is performed, etc. This process is continued until either the first or the second decision is made. The number n of observations required by such a test procedure is a random variable, since it depends on the outcome of the observations.

The problem of sequential analysis arose in this country in the Statistical Research Group of Columbia University about four years ago in connection with a question posed by Captain Schuyler in the Navy Department. M. Friedman and W. Allen Wallis, both members of the Statistical Research Group at that time, recognized the great potentialities of the sequential method of sampling and suspected that a sequential test procedure may exist which requires, on the average, a substantially smaller number of observations than equally reliable tests based on a predetermined number of observations. They proposed the problem of sequential analysis to me and the results of my investigations then led to the development of the so-called sequential probability ratio test [1], [2].*

Independently of the developments in this country, G. A. Barnard in England recognized the merits of sequential sampling. A summary of some of the work he and his associates had done during the war along these lines was recently published [9].

It has been shown (section A.7 in [2]) that the sequential probability ratio test effects the greatest possible saving in the average number of observations, when used for testing a simple hypothesis against a single alternative. It frequently results in a saving of about 50 percent in the number of observations over the most efficient test procedure based on a fixed number of observations.

The sequential probability ratio test for testing a simple hypothesis H_0 against a single alternative hypothesis H_1 is defined as follows: Let $f(x, \theta)$ be the probability density function of the chance variable X under consideration. Let, furthermore, H_0 be the hypothesis that $\theta = \theta_0$ and H_1 the hypothesis that $\theta = \theta_1$ where θ_0 and θ_1 are given values. The probability density of the sample of m observations $x_1, \ldots, x_m$ is then given by

$$P_{1m} = f(x_1, \theta_1) \ldots f(x_m, \theta_1) \text{ when } H_1 \text{ is true}$$

and by

$$P_{0m} = f(x_1, \theta_0) \ldots f(x_m, \theta_0) \text{ when } H_0 \text{ is true.}$$

At each stage of the experiment, at the mth observation for each integral value m, we compute the probability ratio $\frac{P_{1m}}{P_{om}}$. If

$$(1) \qquad \frac{P_{1m}}{P_{0m}} \leqq \text{a certain preassigned value } B$$

---

* Numbers in brackets refer to literature references listed at the end of this paper.

we terminate the process with the acceptance of H_0. If

$$(2) \qquad \frac{P_{1m}}{P_{0m}} \geqq \text{a certain preassigned value } A$$

we terminate the process with the rejection of H_0. If

$$(3) \qquad B < \frac{P_{1m}}{P_{0m}} < A$$

we continue the experiment by taking an additional observation.

For purposes of practical computations it is convenient to take logarithms on both sides of the above inequalities. The reason for this is that $\log \frac{P_{1m}}{P_{0m}}$ can be written as the sum of m terms

$$\log \frac{P_{1m}}{P_{0m}} = \log \frac{f(x_1, \theta_1)}{f(x_1, \theta_0)} + \ldots + \log \frac{f(x_m, \theta_1)}{f(x_m, \theta_0)}$$

which is easier to compute than the original ratio $\frac{P_{1m}}{P_{0m}}$.

In carrying out such a test procedure, we may commit two kinds of errors. We may reject H_0 when H_1 is true (error of the first kind), and we may accept H_0 when H_1 is true (error of the second kind). Let α denote the probability of committing an error of the first kind, and β the probability of committing an error of the second kind. It has been shown that the following inequalities hold :

$$(4) \qquad A \leqq \frac{1-\beta}{\alpha} \quad \text{and} \quad B \geqq \frac{\beta}{1-\alpha}$$

If the alternative hypothesis H_1 is fairly near the null hypothesis, that is, θ_1 is fairly near θ_0, for all practical purposes we may replace the inequality sign in (4) by the equality sign. Thus, if θ_1 is fairly near θ_0 and if we want a test procedure such that the probability of an error of the first kind is equal to α and the probability of an error of the second kind is equal to β, for all practical purposes we may put

$$A = \frac{1-\beta}{\alpha} \quad \text{and} \quad B = \frac{\beta}{1-\alpha}$$

Apart from a slight approximation (caused by neglecting the excess of the probability ratio $\frac{P_{1m}}{P_{0m}}$ over the boundaries A and B at the termination of the process), the sequential probability ratio test is an optimum procedure in the sense that the average number of observations needed for the test is a minimum as compared with any other (sequential or non-sequential) test which controls the two kinds of errors to the same extent.

Although the sequential probability ratio test is defined for testing a simple hypothesis against a single alternative it can be used for testing simple and composite hypotheses concerning θ also when θ can take infinitely many values, as it is usually the case in applications. A general theory for the construction of the sequential probability ratio test in the case when the parameter θ can take infinitely many values, together with a brief discussion of a number of important special cases, is given in [1] and [2]. A more detailed discussion of a number of applications of the sequential probability ratiotest, together with a considerable number of tables, charts, and computational simplifications, is obtained in a report of the Statistical Research Group of Columbia University [3]. Among the special applications discussed in [1], [2], and [3] are, for example, tests relating to the mean of a binomial distribution, tests relating to the mean of a normal distribution when the standard deviation is known or unknown, and tests relating to the standard deviation of a normal distribution when the mean is known or unknown. Further applications relating to the comparison of two populations, as well as interesting theoretical contributions, are contained in a paper by M. A. Girshick [4].

As an illustration, I should like to discuss briefly an example. Let the chance variable X under consideration be normally distributed with unknown mean θ and known variance. Suppose that we wish to test the hypothesis H that $\theta \leqslant \theta'$. It will, in general, be possible to determine two values $\theta_0 < \theta'$ and $\theta_1 > \theta'$ such that the rejection of H is considered an error of practical importance whenever $\theta \leqslant \theta_0$, the acceptance of H is regarded an error of practical importance whenever $\theta \geqslant \theta_1$, while for θ between θ_0 and θ_1 we do not care particularly which decision is made. Under such circumstances we shall want a test procedure for which the probability of rejecting H does not exceed a certain preassigned value α whenever $\theta \leqslant \theta_0$, and the probability of accepting H does not exceed a certain preassigned value β whenever $\theta \geqslant \theta_1$. These conditions will be satisfied by the sequential probability ratio test for testing θ_0 against θ_1, the constants A and B being determined so that the probabilities of the errors of the first and second kinds are equal to α and β, respectively. The sequential probability ratio test leads to a procedure which can be represented graphically as follows: Two parallel lines L_0 and L_1 are drawn whose intercepts and common slope depend only on α, β, θ_0, and θ_1. Thus, these two lines can be constructed before experimentation starts. At each stage of the experiment, at the mth observation for each m, the cumulative sum $\sum_{i=1}^{m} x_i$ of the first m observations is computed and the point $(m, \Sigma x_i)$ with abscissa m and ordinate $\sum_{i=1}^{m} x_i$ is plotted. We continue taking additional observations as long as $(m, \Sigma x_i)$ lies between the two lines. At the first time when $(m, \Sigma x_i)$ does not lie between the two lines the process is terminated. The hypothesis H is accepted if $(m, \Sigma x_i)$ lies on L_0 or below, and H is rejected if $(m, \Sigma x_i)$ lies on L_1 or above.

Practical applications of sequential analysis have been made so far mostly in connection with acceptance inspection and quality control of manufactured products. The application of sequential analysis is, however, by no means restricted to this field. It can be used with advantage in all empirical fields of investigation where the data are not given historically but have to be obtained from experiments which can conveniently be arranged in a time succession so that at each stage of the experiment the experimenter can freely decide whether experimentation be terminated or an additional trial be performed. In some experimentotion the performance of a trial may require a relatively long period of time. In such cases it may not be practicable to wait with the performance of the next trial until the result of the preceding trial is known. Instead, a large number of trials have to be performed simultaneously. Under such conditions sequential analysis is, of course, not practicable.

Sequential analysis has been developed so far mainly for the purpose of testing a statistical hypothesis. A more general problem is this: suppose that k mutually exclusive and exhaustive hypotheses $H_1, \ldots, H_k$ are formulated and the problem is to accept one of them on the basis of the sample. This multi-decision problem contains the problem of testing a hypothesis as a special case when $k = 2$. A possible approach to the problem of sequential multi-decisions is briefly outlined in [2]. This field, however, is not yet sufficiently explored and further progress is still a matter of future development.

Another important problem, about which we do not yet know much, is that of sequential estimation. The problem of sequential interval estimation may be formulated as follows: A class C of intervals is given. For example, the class C may be the class of all intervals of a given length. Furthermore, a number α between 0 and 1 is given. The problem is then to construct a rule which tells the experimenter when to stop experimentation and what element of C be selected at the terminal stage. The rule must be constructed so that the probability that the selected interval will cover the true parameter value is at least α. A general sequential estimation procedure satisfying this condition is briefly described in [2]. It is, however, rather doubtful whether the procedure given there is anywhere near an optimum procedure. The problem of finding a general method for obtaining efficient sequential estimation procedures is still far from solved. For the case of estimating the mean of a normal distribution with unknown variance, C. Stein [5] has devised a very interesting sequential two-sample procedure which leads to confidence intervals of fixed length independent of the value of the population variance.

Finally, I should like to mention some important recent results concerning sequential point estimation. The problem of obtaining unbiased estimates of an unknown parameter on the basis of a sequential sample is by no means a simple one. This problem was solved for the binomial case by Girshick, Mosteller and Savage [6]. More general results were obtained by Blackwell [7]. Important results giving a lower bound for the variance of any sequential estimate were obtained by Wolfowitz [8].

References

[1] A. Wald, "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp. 117-186.

[2] A. Wald, *Sequential Analysis*, (John Wiley & Sons, Inc.,) New York, 1947.

[3] Statistical Research Group, Columbia University, *Sequential Analysis of Statistical Data: Application*, Columbia University Press, New York, 1945.

[4] M. A. Girshick, "Contributions to the Theory of Sequential Analysis," Part I, *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 123-143; Parts II and III, *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 282-298.

[5] C. Stein, " A Two-sample Test for a Linear Hypothesis Whose Power is Independent of the Variance," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp. 243-258.

[6] M. A. Girshick, F. Mosteller, and L. J. Savage, " Unbiased Estimates for Certain Binomial Sampling Problems With Applications," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 13-23.

[7] D. Blackwell, " Conditional Expectation and Unbiased Sequential Estimation," *Annals of Mathematical Statistics*, Vol. 18, (1947), pp. 105-110.

[8] J. Wolfowitz, "The Efficiency of Sequential Estimates and Wald's Equation in Sequential analysis," *Annals of Mathematical Statistics*, June 1947 (in print).

[9] G. A. Barnard, "Sequential Tests in Industrial Statistics," *Supplement to the Journal of the Royal Statistical Society*, Vol. 8 (1946), pp. 1-21.

Résumé

La méthode de Score ("sequential analysis") a été développée, dans les quatre dernières années, spécialement en vue du problème de tester la vraisemblance d'une hypothèse statistique. Une caractéristique essentielle de cette méthode consiste en ce que le nombre des observations demandées pour le test n'est pas fixé à l'avance, mais qu'il dépend des résultats des observations. La procédure est la suivante: Une règle est donnée permettant à chaque stade des opérations (à la m-ième observation pour chaque valeur intégrale de m) de savoir si l'on doit: 1) accepter l'hypothèse H soumise au test; 2) rejeter H; 3) continuer l'opération par une observation supplémentaire.

Il a été montré que le soi-disant "sequential probability ratio test" permet la plus forte réduction possible du nombre moyen d'observations, lorsqu'on l'utilise pour tester une hypothèse simple contre une seule alternative. On obtient souvent une réduction d'environ 50% du nombre des observations sur la procédure la plus efficiente basée sur un nombre fixe d'observations. Des applications du "sequential probability ratio test" ont été faites à plusieurs problèmes spéciaux, tels que des tests concernant la moyenne d'une loi binomiale ou de Poisson, concernant la moyenne d'une loi normale quand l'écart-type est connu ou inconnu, des tests concernant l'écart-type d'uné loi normale quand la moyenne est connue ou inconnue, etc.

Comme il est dit plus haut, la Méthode de Score a été développée principalement en vue de tests d'hypothèses statistiques. Un autre problème important, au sujet duquel on ne sait encore que peu de chose, est celui de l'estimation par la Méthode de Score. D'importants résultats ont été obtenus récemment par divers auteurs américains concernant des estimations, sans erreur systématique, appelées "unbiassed sequential point estimates" et leurs variances.

Discussion

Professor Tornqvist: In recent years the theory of Sampling, Estimation, and Sequential Analysis has made good progress. We all know the remarkable works of Fisher, Neyman, Frisch, Cramér, Wold, Reiersøl, and Haavelmo. We have today heard of many new interesting results obtained by distinguished American and English authors. Also among German authors, I think of H. Gebelein's interesting attempt in his book "Zahl und Wirklichkeit" and of S. Koller's book "Graphische Tafeln zur Beurteilung statistischer Zahlen," which have made valuable contributions to the subject. Two Finnish authors, K. Karhunen and G. Elfving, have recently published works on the theory of sequential analysis. The former's dissertation treatise for the degree of doctor, "Uber lineare Methoden in der Wahrscheinlichkeitsrechnung," contains a strictly mathematical foundation for the theory of linear expressions of stochastic functions. Elfving has published works on stochastic sequences especially on Markoff processes.

The theory of sampling deals with the problem of determining the distribution functions for functions of sampled values, when the distribution function of the parent population is known. The theory of estimation deals with the inverse problem of estimating functions (or parameters) of the parent population or of new samples, by means of a sequence of sampled values. For the case of a finite parent population Hans Gebelein gives in his book some interesting results of which I will speak here. He solves the problem of inversion in a new way.

When we take a random sample containing $[n]$ individuals in all from a parent population with $[N]$ individuals distributed among k alternative classes as the number system $N = (N_1, \ldots, N_k)$ describes, the probability of getting a sample with n_1 individuals belonging to the class n:o 1, n_2 individuals to the class n:o 2, and so on, is, as is well known,

$$(1) \qquad p(n; [n], N) = \binom{N}{n} \Big/ \binom{[N]}{[n]}; \quad \binom{N}{n} = \prod_{i=1}^{k} \binom{N_i}{n_i}.$$

In the case, when the total number of individuals in the parent population $[N]$ and a sample n' are the only facts which are known, Gebelein defines the inverse probability of the possibility that in the parent population there were $N = (N_1, \ldots, N_k)$ individuals in the different classes, by means of the formula

$$(2) \quad p(N; [N], n') = \binom{N}{n'} \Big/ \binom{[N]+k-1}{[n']+k-1}; \binom{[N]+k-1}{[n']+k-1} = \sum_{N;\,[N]} \binom{N}{n'}.$$

The most interesting formula is, however, the formula for the "sequential probability"

$$(3)\quad p(n;[n],n',[N]) = \binom{n+n'}{n'} \Big| \binom{[n]+[n']+k-1}{[n']+k-1}.$$

This sequential probability will numerically fix our anticipations of getting a new sample, in which the individuals are distributed among the k different classes as the number system $n = (n_1, \ldots, n_k)$ describes, from the same parent population from which we already have taken a sample with individuals distributed as the number system $n' = (n_1', \ldots, n_k')$.

The probability distribution (3) for the sequential probabilities is consistent with (1) and (2) because we have in fact

$$p(n;[n],n,[N]) = \sum_{N,[N]} p(N;[N],n')\, p(n;[n],N-n').$$

The sequential probability concept is thus independent of $[N]$.

The weakest point in Gebelein's theory is, however, the concept of the number of classes k of the parent population. He seems to have made an implicit assumption that in every class taken into consideration there is at least one individual present in the parent population. In the cases when the classes are got by subdivisions of larger intervals this need not always be true. The number k can then be arbitrarily chosen by the investigator. It is, however, interesting to study the suggested probability distribution function for the sequential probabilities.

Gebelein meets very great mathematical difficulties when trying to calculate moments of linear functions depending on the probability distribution studied.[1] By means of probability functionals this may, however, be done rather easily.

The probability functionals corresponding to the three probability distributions (1), (2), and (3) can be expressed by the formulas (4), (5), and (6).

The random sampling probability functional is:

$$(4)\quad \overset{o}{\underset{y}{\uparrow}} \prod_{i=1}^{k} (1+\uparrow_y \uparrow_{n_i})^{N_i} \binom{[N]}{y}^{-1} \phi(y;[n]) = G(\uparrow_n;[n],N)$$

The probability functional for Gebelein's inverse probability distribution is

$$(5)\quad \overset{o}{\underset{X}{\uparrow}} \prod_{i=1}^{k} (\uparrow_X \uparrow_{N_i})^{n_i'} (1-\uparrow_X \uparrow_{N_i})^{-(n_i'+1)} \binom{X-1}{[n']+k-1}^{-1} \phi(X;[N]) = G(\uparrow_N;[N],n').$$

---

[1] Gebelein studies in full only the case k = 2 and moments of at most second degree.

The sequential probability functional is

$$(6)\quad \overset{o}{\underset{x}{\uparrow}} \prod_{i=1}^{k} \big(1-\underset{x}{\uparrow}\ \underset{n_i}{\uparrow}\big)^{-(n_i'+1)} \binom{X+[n']+k-1}{[n']+k-1}^{-1} \phi(X;[n]) = G(\underset{n}{\uparrow}\ ;[n],n')$$

In these formulas the operator symbol $\overset{n'}{\underset{n}{\uparrow}}$ has the property

$$\overset{n'}{\underset{n}{\uparrow}}\ u(n) = u(n+n')$$

The operator acts in a similar manner on the variable y and so on. The "selecting conditional probability function":

$$\phi(x;A) = \begin{cases} 1 & \text{if } x \subset A, \\ 0 & \text{,, } x \subset A. \end{cases}$$

The mathematical expectation of functions $u(n)$ of the number system n, can formally and often also after some calculations explicitly be calculated by a symbolic multiplication with the corresponding probability functional. The probability that $u(n)$ belongs to a class $[u(n)\subset A]$ can in this way be calculated as the mathematical expectation of the selecting probability function $\phi(u(n);A)$ which takes the value 1 if $u(n)$ belongs to the class A, and the value 0 if $u(n)$ does not belong to this class. The moments of the stochastic variable $u(n)$ in respect to the sequential probability functional $G(\uparrow_n;[n],n')$ can be calculated by a serial development of the symbolical product

$$G(\underset{n}{\uparrow}\ ;[n],n')\,e^{\lambda u(n)} = \sum_{\nu=0}^{\infty} G(\underset{n}{\uparrow};[n],n')\,(u(n))^{\nu}\,\frac{\lambda^{\nu}}{\nu!}$$

according to increasing powers of λ. In the case when $u(n) = [un] = \Sigma u_i n_i$ we can by suitable transformations of the symbolic product so obtained rather easily calculate explicitly the three first moments about the mean

$$G(\uparrow_n\ ;[n],n')\,[un] = [n].\frac{[\bar{N}u]}{[\bar{N}]}\ ;\ \bar{N}_i = n_i'+1$$

The variance will be

$$\sigma^2([un];[n],n') = [n]\ \frac{[\bar{N}]+[n]}{[\bar{N}]+1}\cdot\frac{[\bar{N}(u-[\bar{N}u]/[\bar{N}])^2]}{[\bar{N}]}$$

and the third moment about the mean:

$$\zeta^3([un]; [n], n') = [n]\frac{[\bar N] + [n]}{[\bar N] + 1}\cdot\frac{[\bar N] + 2[n]}{[\bar N] + 2}\cdot\frac{[\bar N(u - [\bar N u]/[N])^3]}{[\bar N]}$$

The corresponding moments for the direct probability distribution $p(n; [n], N)$ are obtained from this formula if we put $N = -N$. The moments (higher than the first) of the inverse probability distribution (2) about the mean

$$G(\underset{N}{\uparrow}\ ; [N], n')\,[un] = [n'u] + [N - n']\frac{[\bar N u]}{[N]}$$

are the same for the sequential probabilities $p(N-n'; [N-n'], n')$.

The distribution functions studied above are all special cases of a great class of distribution functions defined by help of this formula

$$P_r\,(X\subset AB; x\subset B) = \underset{X}{\overset{0}{\uparrow}}\frac{\prod_{i=1}^{k}\left(1-\underset{X_{0i}}{\uparrow}(1+\underset{X_{1i}}{\uparrow}+\underset{X_{2i}}{\uparrow}+\dots)\right)^{-1}\phi(X;A)\phi(X;B)}{\prod_{i=1}^{k}\left(1-\underset{X_{0i}}{\uparrow}(1+\underset{X_{1i}}{\uparrow}+\underset{X_{2i}}{\uparrow}+\dots)\right)^{-1}\phi(X;B)}$$

Functionals of the above (and suitably generalized) forms can successfully be used in attempts to solve other intricate problems of mathematical statistics especially in the theory of transforming time series to new time series which can be described as sampled values from a parent population with a distribution function of some simple form. I will not, however, trouble you longer with these intricate problems. I wish to thank Professor Yates, Professor Cochran, and Professor Wald for their very interesting papers.

Mr. W. R. Thompson:

Mr. Thompson called the attention of the meeting to two early articles on the notions of sequential analysis and associated apportionment.[1]

Dr. Yates:

Professor Cochran has raised one or two points on which I should like to comment. These were not dealt with in my paper because, as I explained, I was asked to give an historical paper, excluding recent developments in the United States, which were to be dealt with by Professor Cochran.

---

[1] William R. Thompson, "On the Likelihood that One Unknown Probability Exceeds Another in view of the Evidence of Two Samples," *Biometrika*, Volume 25 (1933), p. 285; and "On the Theory of Apportionment," *American Journal of Mathematics*, Vol. 57 ((1935).

I did include some reference to work done during and after the war in England but I omitted the details because the principal modern developments have, I think, occurred in this country. The fact that so much has been done in recent years in the United States is an excellent illustration of my thesis that theory always develops in conjunction with practice. In this country you have employed sampling technique to a considerably greater extent than we have, partly, I think, because you are less conservative and partly because you have a larger and more dispersed population. For that reason recent developments in theory have been particularly prolific on this side of the Atlantic.

Professor Cochran mentioned the non-response problem. We have a sub-sample of the non-responding households for further visits in order to cut down the number of these visits, though we haven't developed the theory.

Professor Cochran also mentioned systematic sampling. I might amplify the remarks I made on my own work as they tie in directly with his. Firstly, the question of what I called in my paper "end-corrections." Professor Cochran pointed out that if you have a trend and take a systematic sample this is likely to be more inaccurate than the corresponding stratified random sample. End-corrections are designed to overcome this and consist of altering the weights of the sampling units at the beginning and end of the sequence. Thus, if the first sampling unit is right at the beginning of the sequence it will be given half weight, if it is distant one sampling interval from the beginning of the sequence it will be given a weight of 1½. This type of correction is of value not only in the case of a simple linear trend but in cases where the beginning and end of the sequence are widely different in value.

Secondly, on the question of estimation of the sampling error of systematic samples, I would repeat the conclusion that I have reached, namely that there appears to be no certain way of estimating the systematic sampling error unless additional observations are taken intermediate between the main sampling points. In dealing with a standard type of material it may well be possible, after a thorough investigation, to invent an empirical formula which will give an appropriate estimate of the sampling error from the variation between the different sampling units, but I think one must be very careful not to generalize as each type of material has its own peculiarities. The difficulties will be apparent if one considers the example that Professor Cochran took in his paper of 1946. In that paper he showed that with an autoregressive sequence *i.e.*, with an exponential correlogram, the sampling error could be estimated from the variation between the aggregate of the sampling units and the correlation between neighboring units. If, however, we have autoregressive sequence on which a random variation is superimposed, each unit varying independently, the formula Professor Cochran developed will give a very misleading answer.

Finally, on the problem of sequential sampling I would like to make one point. Sequential sampling can be looked at in two ways; firstly, from the point of view of testing hypotheses, which Professor Wald explained; secondly, like all sampling, from the point of view of estimation of the parameters of the population. In sequential sampling we get the maximum efficiency of estimation when the population parameter has a value corresponding to the H_0 hypothesis, with a corresponding loss of efficiency when the population parameter differs from this value. In industrial testing and similar work where one is concerned with conformity to a certain standard this property is exceedingly useful and results in large gains in efficiency in the sampling process. In sociological and economic sampling, on the other hand, one is generally concerned with the estimation of population parameters, and about the same accuracy of estimation is required regardless of what value the population parameter actually has. For that reason sequential sampling is not applicable to problems of economic and sociological sampling to anything like the same degree as it is to problem of industrial sampling. I make this comment not in any way to deprecate sequential sampling, which I have a very high regard for, but merely in order that it should not be applied where it is not really suitable.

Professor Cochran:

I would like to express my agreement with Professor Yates' views on the question of estimating the error of a systematic sample from the results of a single sample; and to mention that the results which Osborne obtained and which showed nice agreement between his formula and the data, were based on the assumption of an exponential correlogram. He had found that an exponential correlogram fitted the data that he presented extremely well. So these results do assume, as in some degree one must assume, a particular type of auto-correlation structure in the population.

Professor Wald:

I would like to make just a brief comment on the questions raised by Professor Yates. I believe (I don't know whether I understood him correctly) that his comment refers to the actual sampling procedure as in use now, and the theory of which has been developed, namely the sequential probability-ratio test procedure.[*] I would like to mention that it is conceivable that other sequential test procedures, or sampling procedures, can be devised which will give more equal efficiency over the whole scale, and will work from that point of view more satisfactorily than the sequential probability-ratio type of sampling. I wonder (could Professor Yates correct me?) whether his remark referred to the procedures in actual use, or whether he meant that in general no sampling method can be devised which would give good efficiency over the whole parameter space.

Professor Yates:

I am not an expert on sequential sampling, but my own view, for what it is worth, is that if you work out a sampling procedure on the sequential basis which is equally efficient over the whole range you will probably find that it is necessary to take the same number of sampling units, whatever the results obtained in the first *k* sampling units. In other words, the ordinary procedure of taking a fixed number of sampling units will result.

Professor Wald:

I think this has been proved for the normal case; Mr. Stein and myself investigated the case of estimating the mean of a normal distribution. We found that if we want equal efficiency over the whole scale we have to take a fixed-size sample; we cannot do any better. That paper will come out, probably, in the *Annals of Mathematical Statistics*.[1] However, in other types of non-normal distribution, we believe that there is a considerable gain in the sequential method of sampling, even if we merely want estimates which have about equal efficiency everywhere.

---

[1] Volume 18 (1947).

SAMPLING PRACTICE

*Saturday, September 13, at 9:30 a.m.**

CHAIRMAN :
Helen M. Walker

HISTORY OF THE USES OF MODERN SAMPLING PROCEDURES

by Frederick F. Stephan**
Professor of Social Statistics
Princeton University (United States)

I. INTRODUCTION

Modern sampling practice is the result of combining statistical theory with scientific knowledge about the material that is being sampled and with experience that has been gained through the use of various sampling techniques in surveys and experimental work. Its history stems from many roots and its applications branch out into many fields of science, commerce, manufacturing, agriculture, education, and government administration. In spite of its wide range of usefulness, sampling practice has been neglected in the training of statisticians, in the textbooks and treatises, and in the planning and analysis of most experiments and studies [1].[1] However, like Cinderella, it has risen from neglect to a position of well-deserved importance.

---

* Joint meeting with The Econometric Society (arranged by the American Statistical Association).

** This paper was prepared in connection with studies of sampling under the Committee on the Measurement of Opinion, Attitudes and Consumer Wants of the National Research Council and Social Science Research Council.

[1] Figures in square brackets refer to notes and references listed at the end of this paper.

11

The history of sampling practice provides a useful background for the discussion of present theory and applications. Difficulties beset any one who attempts to trace the developing uses of sampling because they are scattered throughout many branches of science and technology and are described, if at all, in subordinate portions of reports and articles whose titles provide no hint of what they may contain on the subject of sampling. Hence, this paper not only will touch each development very briefly and inadequately but no doubt it will miss entirely a number of highly important facts. (The author will appreciate suggestions for correction and addition). The principal emphasis will be on the steps by which the present stage of progress in applied sampling was reached and no attempt will be made so present a comprehensive picture of present-day uses or to appraise the relative value of each contribution to the accumulation of technical knowledge.

There are two dangers that both author and reader must avoid in an historical review: (a) they may read into the records more than was actually there, and (b) they may assume that nothing existed that is not given in the records. I hold to the view that the principles and technical practices of sampling emerged rather gradually from notions that were simpler, sometimes inaccurate and confused, and often not recognized by those who held them as being of any special importance. Hence the records are fragmentary and there were doubtless many instances in which principles were developed by two or more workers in ignorance of the other's work. I would not say that the same discovery was made "independently" for there is always a flow of common ideas, sometimes rapid and sometimes slow, in which all share. Hence it should be understood that the particular examples that will be mentioned are part of a larger development of practice and theory which connected them in various known and unknown ways and often influenced profoundly their characteristics and effectiveness.

Finally, greater attention will be given to the use of sampling in large-scale surveys rather than to its use in the laboratory or in small-scale (local) experiments. This accords with the emphasis of the other speakers who will follow.

II. EARLY EXAMPLES OF THE USE OF SAMPLES

The earliest examples of sampling procedures are to be found in certain very ordinary human activities. The common practice of taking a small part or portion for tasting or testing to determine the characteristics of the whole precedes recorded history and is one of the roots from which sampling methodology stems. Stirring and mixing before taking the sample is a prototype of randomization. The efforts of scientists and their predecessors to draw conclusions about the laws of nature from what they could observe in their immediate environment was a sampling process. Astronomy, perhaps the oldest of the sciences, started with the moon, the larger planets, and the stars that one could see from positions near the equator or north of it. Even today it is limited to those heavenly bodies that can be photographed with the best available instruments.

All empirical knowledge is, in a fundamental sense, derived from incomplete or imperfect observation and is, therefore, a sampling of experience. An unusually interesting example, not unrelated to modern population research, can be found in Halley's selection of mortality statistics in Breslau to form the basis of his life table. From this sample of one city he drew general conclusions pertaining to "the mortality of mankind." [2]. This was in 1693. Two centuries later, Sir John Lawes likewise used the annual record of wheat yields at Rothamsted on 5 plots totalling 33 acres to estimate the change in yield per acre from 1852 to 1879 for all of England and Wales [3]. While these examples were both statistical, all scientific observation, whether statistical or not, is based on sampling.

In spite of the widespread practice of sampling in commerce and industry and its utilization in scientific research, official statisticians have usually set as their goal the complete counting of a population and have sought to avoid any use of sampling or any inference beyond the bare description of that population at the time it was counted. Still they have been compelled to compromise in the face of practical difficulties, on something less than complete coverage and often also on something less than adequate accuracy and quality in the information recorded. Such compromises have been common in the collection of many kinds of statistics, particularly those usually collected by voluntary reporting such as employment, wage rates, prices, and crop yields. It is also found in the collection of data that ought to be complete in order to satisfy certain non-statistical purposes, as for example of the registration of births. Often the official statistician, faced with great difficulties, opposition, and indifference on the part of the public and with only weak powers to support his efforts, could only compromise or resign his post. The result of the compromise was often a very crude and unsatisfactory form of sampling which strove to get as much information "as possible" and accepted what it got, instead of attempting to control the selection by some of the means that were available.

Interestingly enough, sampling has generally preceded the establishment of regular censuses. Prior to first British Census in 1801 the size of the population of England was not known very accurately and various speculations were made and estimates prepared and disputed.

Malthus projected his famous theory of population, and Adam Smith his *Wealth of Nations*, before the establishment of periodic national censuses. At least as early as 1754, estimates of population of England were made from the number of houses on the tax list plus a rude estimate of cottages not taxed, the total of dwellings being multiplied by a somewhat arbitrary factor of 6 persons per dwelling [4]. Other estimates were based on the reported number of baptisms, marriages, and burials.

In 1800 Sir-Frederick Morton Eden estimated the population of Great Britain at 9,000,000 using sample data on the average number of inhabitants per house as well as the number of births. The first census of Great Britain in 1801 confirmed his estimate. Somewhat similar methods of estimation have been used between censuses in the U.S.A. as recently as the last decade [5].

About 1765 Messance, and 1778 Moheau, published very carefully prepared estimates for France based on enumeration of population in certain districts and on the count of births, deaths, and marriages as reported for the whole country. The districts from which the ratio of inhabitants to births was determined truly constituted a sample. Laplace prepared similar estimates in 1802 from enumeration of 30 departments of France and the reports of births, deaths, and marriages from 1799 to 1802. This followed a plan he published in 1786. He made a remarkable step forward in attempting to measure the precision of his estimate and announced that the odds were 1161 to 1 that it was not in error by more than 500,000 inhabitants or 2 per cent of the total population. Even though the method of estimate was crude and the measure of precision not wholly valid, Laplace's effort was much more successful than the complete census of France that was attempted at the same time. A similar procedure had been employed in 1784-91 by the famous chemist Lavoisier to estimate the number of horses, cattle, sheep, and pigs, as well as the area under cultivation [6].

The foregoing examples suggest that modern sampling procedure might have developed at least a century sooner than it did if it had received more attention from the scientists of the day. Only the officials of statistical bureaus and other posts continued to concern themselves with the problems of measuring population, and they were preoccupied with difficulties in the classification and interpretation of data, in the administration of statistical bureaus, and in the important problems of trade, finance, industry, agriculture, public health, etc., for which statistical data were needed. Hence they favored complete censuses or the closest approach to them that was feasible.

III. Examples of Development of Need for Efficient Sampling Methods

It would be interesting to explore the development of statistical work in its many ramifications and inquire about the places in which modern sampling methods could have been used to good advantage. That is beyond the possibilities of a paper such as this. Instead examples will be given from the beginnings of four general lines of statistical work in which efficient sampling methods were needed: (a) agriculture crop estimates, (b) economic statistics of prices, wages, employment, etc. (c) statistical phases of social surveys and health studies, and (d) public opinion polling. The more recent developments of the use of sampling in these and other fields will be traced in succeeding sections.

(*a*) *Agricultural Crop and Livestock Estimates:* At the present time a tremendous variety of statistics are collected in the United States on acreage planted in each principal or special crop, the estimated yield as judged at successive dates during the growing season, actual yields, numbers of livestock, equipment, farm labor, marketing, and other aspects of agricultural production. Other countries collect similar statistics with variations in scope, detail, and frequency. In the United States, as in many countries, the methods used for this purpose include periodic censuses, special surveys, voluntary reporting by selected respondents, and records produced in connection with taxation, marketing, and foreign trade.

This vast system had its beginnings about a century ago. The first U.S. Census of Agriculture was taken in 1840. At the same time collection of agricultural statistics was begun by the Patent Office. Monthly reports were collected and published by the editor of the American Agriculturist in 1862 and in 1866 regular reports on acreage, condition of crops, yield, and livestock were begun in the newly established Department of Agriculture. Annual reports of prices were added in 1867. Thereafter the system developed into an extensive organization of agents and voluntary reporters that increased from 4,000 in the late 70's to 300,000 in 1931 and a somewhat larger number now [7]. (The present system is described in a paper presented by Dr. C. F. Sarle at this session). The methods of sampling were quite crude, many of the reports were simply based on the respondent's judgment about conditions in his locality and the list of respondents was built up of farmers, agricultural agents, members of the staff of agricultural colleges, etc., who were willing to serve without pay and were located at places widely scattered over the country. The system was a compromise between the excessive cost and slow returns of complete enumeration on the one hand and the wholly unsatisfactory alternative of relying on reports issued by private speculators or going without current information, on the other. Various attempts were made to improve the reliability of the statistics that resulted from these reports but they continued to be biased and of limited accuracy. Nevertheless they have been of great value to farmers and to other users of the statistics and they are the principal source of current data on agriculture. Here as elsewhere the introduction of modern sampling methods came late but is now making great contributions to the improvement of the statistical information in this field.

(*b*) *Economic Statistics:* The history of economic statistics is a long and detailed story which can only be sketched here in inadequate outline [8]. Again, as for population and agriculture, there were early attempts to make the best of scanty and fragmentary material. Censuses and special investigations have provided increasingly valuable data but at long intervals and in inadequate detail.

In many countries, economic statistics derived primarily from the records produced by taxation and customs duties. This was true in the United States. Statistics about subjects that were not provided by the system of taxation had to be obtained by other means. More than a century ago the collection of reports of prices, wage rates, hours of work, employment, and production was begun with voluntary reporting by employers and occasional special surveys as the principal methods. Massachusetts established a statistical bureau in 1869, Pennsylvania in 1870, and other states at later dates; the Federal Government set up a Labor Bureau in 1884. These bureaus developed staffs of field investigators to collect data at monthly or other periods but relied heavily on mail questionnaires for many of their principal series. This continues to be the prevalent procedure although the establishment of social insurance systems and employment exchanges has provided other sources of statistics on earnings and employment.

Similar systems of collecting reports are operated by other agencies concerned with production, trade, and finance. Those inquiries have been based on simple sampling procedures such as the use of a scattered group of voluntary reporters, except for certain relatively recent surveys. A serious effort was made, for example, to get wholesale prices "in representative markets" and to get labor data by sending agents into various districts with a list of employers from which they could choose what they believed to be a representative group but the techniques of selecting a representative sample had not been developed as formal procedures. Here again modern sampling procedures have been introduced late but are making noteworthy contributions, in conjunction with other improvements, to the accuracy and value of these statistics.

(c) *Social Surveys and Health Surveys:* The distinction between social surveys and certain related kinds of inquiries is difficult to make. Some, like Le Play's monographic studies, are very intensive observations of a few individuals or families; others include thousands of cases but less intensively and with more attention to statistical analysis. Williams and Zimmerman listed 1500 studies of family living in the United States and elsewhere, made prior to 1935 [9]. Booth's *Survey of London Life and Labour,* Rowntree's survey of poverty in York, the reports of the Immigration Commission in 1907, and the Pittsburg Survey of the Russell Sage Foundation are examples. While complete and thorough investigation of the problems was the ideal to which such surveys aspired, they were necessarily limited to a selected city, and to somewhat fragmentary data on pertinent questions by the sheer impossibility of a comprehensive canvass. Other surveys were made under the inspiration of their example but progress in methodology came slowly and as in many other investigations the accuracy of the results was seldom tested.

Sampling has been used extensively in studies of poverty and unemployment. During the depression of 1873-79 Carroll D. Wright used police in 19 cities and wrote to assessors in 375 towns throughout Massachusetts inquiring about the number of unemployed. His estimate of 28,508 corrected the current assertions that 200,000 or 300,000 workers were out of work [10]. Similarly in 1873 the number of unemployed in New York City was estimated on the basis of reports from 400 volunteer visitors of the Association for Improving the Condition of the Poor. These procedures were, of course, crude and inaccurate but they reflected a praiseworthy desire to measure what was being discussed in still less accurate terms by important public leaders. In 1893, unemployment was estimated indirectly by the decline in factory employment from its highest figure in a series derived from a sample of reporting employers constituting 70 percent of the production in the state [11]. In 1893, estimates were collected by mail from respondents in 119 cities by Bradstreet. In New York City, Chicago and other cities, police and health inspectors surveyed samples of homes or factories [12]. In the depression of 1914-15, many surveys of unemployment were made, notably by canvass of the workers insured by the Metropolitan Life Insurance Co. through its agents in 29 cities [13] and by canvass of 104 city blocks and 3703 additional tenement houses by the Bureau of Labor Statistics using about 100 tenement house inspectors. Both agencies repeated their surveys in New York City eight months later [14].

In the depression of 1921-22, somewhat similar methods were used to estimate unemployment and, in general, while the results were far from satisfactory, there was improvement in the identification of the unemployed as distinct from other needy groups and increased detail in the information collected. Especially noteworthy as an advance beyond previous practice was the survey conducted in Columbus, Ohio, by Frederick E. Croxton and Mary Louise Mark using 100 university students to canvass three districts within the city selected by the presidents of the Federation of Labor and the Chamber of Commerce as a "fair sample of the wage-earning population of the city" and including approximately 10 percent of the wage-earning population. The survey was repeated annually from 1922 to 1925 [15]. A similar procedure was used in Philadelphia by J. Frederic Dewhurst and Ernest A. Tupper who had school attendance officers canvass 166 small districts selected to provide an accurate cross-section of unemployment [16]. This survey was repeated annually during the 30's and provided several kinds of data not previously available. The Columbus procedure was used in 1929, 1930, and 1931 in Buffalo, N.Y. by Croxton [17] and in 1931 in Syracuse, N.Y. by John N. Webb [18].

Among numerous health surveys made in the U.S. from time to time, several examples may be mentioned that resemble the preceding surveys in their methods of sampling. In 1921, the U.S. Public Health Survey selected Hagerstown, Md., as a conveniently located and "fairly typical small city in the east not greatly influenced by immigration" in which to study sickness and fertility by periodic visits to a sample of households [19]. In 1929-30 the Committee on Costs of Medical Care made many sample studies, including one on the incidence of illness among 9,000 families [20]. The U.S. Public Health Service and Milbank Memorial Fund studied sickness among those elements of the population that had borne the brunt of the depression by selecting severely effected districts in 10 localities, omitting the slums, Negro, and well-to-do sections [21].

(*d*) *Public Opinion Polls:* The practice of surveying public opinion emerged from simple but obscure beginnings in the "straw vote" conducted by newspapers to sense public reactions to candidates and obtain human interest stories by interviewing the "man in the street." Even before 1900 the New York *Herald* collected pre-election reports and estimates from all over the United States and attempted to forecast the outcome of elections. In 1904 it polled 30,000 registered voters in New York City. In 1905 the Chicago *American* and Chicago *Journal* took a poll during the mayoralty campaign and the Columbus *Dispatch* started a long series of polls by taking one in the state election of 1906. In 1912 papers in Boston, Chicago, Cincinnati, Denver, Los Angeles, and St. Louis conducted a nation wide poll in 37 states. Four years later, this group of papers repeated its presidential election poll and another poll was taken by the Hearst newspapers. The Rexall drug store chain took a nation wide poll in 1920 using its stores as centers to collect the ballots. These and other efforts were outdone in size by the *Literary Digest* which started polling in 1916 and in 1920 mailed out 11,000,000 ballots to persons whose names were taken from telephone directories and other lists. Two years later it conducted a poll on prohibition. Other polls were taken at the rate of about one every year and a half. Robinson found that more than 86 straw polls were taken in the presidential campaign of 1928. Although these polls employed quite crude methods, they were successful in many instances because a moderate error only affected the prediction of the winning candidate when the election was close and because among many attempts some might be expected to turn out well. The sources of error in straw polls were analyzed in detail by Robinson in 1932 [22].

Closely related to the polls were market research studies and consumer surveys conducted by business concerns, publications, and advertising agencies. The results and methods of these studies received less publicity than the polls but they offered even greater opportunities for sampling than the straw votes. There was also a growing interest in public opinion research among political scientists, sociologists, and others which led to scientific interest in the improvement of the technique of opinion polling [23].

The foregoing discussion was limited to uses in which the sampling methods were quite simple, indeed often crude and inefficient, and not based on explicit considerations of probability. The dividing line is not sharp between them and some of the other surveys that will be included in later sections but they represent fairly well the pioneer stage in the emergence of modern sampling practice. This stage might also be illustrated by examples of experiments, surveys, and studies in other fields in which sampling in some form was used.

While, in these early instances, the sampling procedures were simple and usually employed uncritically with no great attention to accuracy and representativeness, it should be noted that the problems of observing and recording data were almost always far more serious than the problems of sampling. Modern sampling procedures, had they been available and had they been applied effectively, would have permitted the use of smaller samples and made possible in most cases more careful fieldwork. Through these developments and those in other fields of statistical study, there arose a great opportunity for the use of better sampling procedures that has been exploited only in part up to the present time.

IV. DEVELOPMENT OF THE USE OF THE SIMPLER RANDOM AND SYSTEMATIC SAMPLING TECHNIQUES

Much of modern sampling practice rests on processes of selecting individuals at random or according to certain systematic procedures. These processes may be employed as providing a simple method for selecting a sample or they may be used as parts of a more complex sampling scheme.

Random sampling utilizes the devices of games of chance or other methods that assure to each unit in the source from which the sample is drawn, an equal probability before the draw that it will be included in the sample. It has the advantage of facilitating certain applications of the theory of probability to the analysis and use of the results.

Although the theory of probability was well established in the eighteenth century, its applications to the practical drawing of samples were delayed until the twentieth. Applications to games of chance and lotteries can be excluded from consideration since they were not concerned with estimating the characteristics of a population from a sample. Likewise, applications to the interpretation of data as having been produced by a random sampling process from a hypothetical universe can be excluded since they did not involve the actual practice of drawing samples. Also trials with dice or cards designed to "verify" known probability laws or merely for expository purposes are not examples of sampling practice. Doubtless there are many instances that have escaped the author's notice, but the earliest instances he has found are Bowley's sampling of a list of bonds and their interest rates in 1906 using the final digits in one of the tables in the *Nautical Almanac* to make the random selection and Student's testing of the *t*-distribution in 1907 by drawing cards from a large receptacle. Random sampling was facilitated by Tippett's tables of random numbers, published in 1927 [24]. The practical importance of random selection slowly gained recognition, especially in connection with more complex procedures that were developed after 1920 and will be discussed in the next section.

Systematic sampling employs a simple rule of counting cases in some convenient order and selecting for the sample every *n*'th case, or of using some similar pattern of selection such as taking material at measured intervals or drilling samples out of a large mass of metal by use of a template that locates the holes. More complex rules or patterns may be used and the starting point may be selected by a random choice or special rule. The theory of probability can also be applied to the results of systematic sampling under certain conditions. Bowley and others have interpreted it as a case of stratified random sampling and estimated sampling errors accordingly.

Systematic sampling has long been in use in taking samples of ore from the face of a vein in a mine or samples of metal from a pig or ingot. The practice of "cruising" a forest and sampling trees at uniform intervals has been well established in forestry for many years [25].

Systematic selection was utilized by A. N. Kiaer in his survey of Norwegian workers in 1895. As a means of facilitating special tabulations from census schedules, it was used in a study of family data in the 1900 Norwegian census, a study of marriage in the Danish census of 1901, and a housing study for Oslo in 1913-14 [26].

Kiaer's work apparently had little influence on surveys in the U.S. or in Europe. Apart from the discussions he initiated in the *Bulletin* of the International Statistical Institute from 1896 to 1903, it was mentioned only briefly by Edgeworth in his Presidential address before the Royal Statistical Society in 1912 and in the U.S. by W. B. Bailey and F. S. Chapin in 1906 and 1920 respectively [27]. It had a greater effect in certain continental surveys [28].

In 1909 Bowley included a brief chapter on Sampling in his Elementary Manual of Statistics making general references to sampling in commerce, mining, and industry, but not to specific sample surveys. However, in 1912 he took every twentieth working class household in Reading in a study of poverty and his associates did the same for three other cities in 1913 [29]. He computed probable errors of sampling and recognized other sources of error in the interviewing, definitions, and process of estimating. This study was followed by a systematic sampling of census schedules in 1915, and by similar surveys of Liverpool (1930) and Merseyside (1931) by Caradog Jones, Southampton (1927*f*) by Ford, and London (1929) by the London School of Economics under Bowley's direction [30].

John Hilton made a series of studies of workers in the unemployment insurance system beginning in 1923 [31]. They were selected systematically from the files of the Labour Exchanges. In addition to the information in the records some data were obtained by interviewing workers who came to the exchange. Certain deviations from strictly systematic selection introduced biases which were reduced by subsequent improvements of the method. Hilton found a sample of only one percent quite satisfactory to meet the practical administrative and policy-making purposes for which the studies were made. The reduction in expense that resulted from sampling such a small proportion of the records was indeed impressive. Oddly enough the method was not imitated to any great extent by other government bureaus.

Another example of the great usefulness of sampling to get quick inexpensive tabulations was furnished by the Japanese who took a sample of one in 1,000 from their 1920 census to get data quickly after the earthquake of 1923 [32]. Subsequent checking when the usual tabulations were completed verified the accuracy of the sample results. Further examples of sampling census schedules for special tabulations are Day Monroe's study of Chicago families in the 1920 Census and R.F. George's study of British workers in the 1931 Census [33].

In spite of this successful experience, the use of random and systematic sampling procedures in statistical work made slow progress. At the Rome Session of the International Statistical Institute in 1925 a resolution was adopted recommending the use of sampling for statistical purposes with appropriate precautions as to its representativeness, mathematical statement of the precision, and full description of the methods employed. The reports that were submitted by the commission that drafted the resolution presented evidence of the usefulness of sound sampling procedures but both the reports and the resolution had less immediate effect on statistical practices than might have been expected. Eight years later, in 1934, Neyman revived the discussion of these reports in a notable paper before the Royal Statistical Society.

In America, serious attention was given to problems of sampling methodology by several committees of the Social Science Research Council. There was a strong current of interest among the sociologists including Chaddock, Ross, Ogburn, Stouffer, Lundberg, Dorothy Thomas, Stephan and others [34]. Margaret H. Hogg, who had worked under Bowley's direction on some of the British surveys, came to America to the staff of the Russell Sage Foundation and there made a critical study of employment and unemployment statistics. In an article in the *Journal of the American Statistical Association* she made a strong plea for rigorous methods of sampling and cast doubt on the value of surveys such as those that had been made in Philadelphia and Buffalo, in which the sample was selected by judgment rather than random procedures. She was equally concerned about the adequacy of the classification of workers into the various categories of employment and unemployment and the analysis of the data in terms of significant questions. In the spring of 1931 Miss Hogg made a survey of unemployment in New Haven, Connecticut, partly for the purpose of testing the practical difficulties of applying a random sampling method and also developing better schedules and statistical categories for unemployment surveys [35].

Two very important general developments affecting the use of sampling in the United States occurred in 1933. One was organization of large scale work projects for the unemployed under the national programs of the Federal Emergency Relief Administration (1933-35), Civil Works Administration (1934-35), and Work Projects Administration (1935-40). The second was the enlistment of many leading statisticians from the universities and business in the reorganization of government statistical work and administration of emergency agencies [36]. The statistical needs of the government increased tremendously as it took active steps to meet the problems of the depression and undertook various New Deal programs. Since the American statistical system is decentralized, it became necessary to establish a Central Statistical Board to coordinate and regulate the statistical activities of many agencies engaged in collecting data for their own purposes or for general use. The Board was largely an outgrowth of the Committee on Government Statistics and Information Services, an advisory group organized by the Social Science Research Council and the American Statistical Association at the request of the Secretaries of Agriculture, Commerce, Interior, and Labor to assist them in overhauling the statistical work of their Departments.

The great variety of statistical work done by the regular agencies, the National Recovery Administration, and other emergency agencies far exceeds the scope of this paper. Only a few large scale surveys can be described briefly as examples of those of their projects that were based on sampling; many smaller and more specialized studies were also undertaken. A number of difficulties prevented these projects from embodying fully the best sampling procedures then known and from exhibiting the range of uses to which sampling methods could be put. Most of the work was started with inadequate time for planning and preliminary trials. There were, curiously enough in a period of unemployment, severe shortages of persons technically competent to supervise the work. Moreover the field workers and office staffs were drawn largely from the unemployment relief lists and worked under restrictions as to hours and rates of pay. In spite of these difficulties the results contributed greatly to many very important programs and provided detailed statistical information where there had been little or none before.

(1) The Financial Survey of Urban Housing was taken in conjunction with the Real Property Inventory, a complete canvass of housing in May 1934 covering 60 cities. The Financial Survey was taken by interviewers who visited all the families in one block out of every ten in larger cities and one in seven in smaller cities. Additional schedules were obtained by mail from families in another four blocks out of ten. The survey resulted in a 1200 page report of detailed financial data and was unique as an early instance of the use of sampling procedure for obtaining special information as part of a single complete canvass [37].

(2) A series of research studies of workers on relief and related subjects were undertaken by the FERA and WPA in selected cities and counties from 1933 to 1940 using a variety of methods and samples.

The localities were selected in part by judgment of their representativeness of different types of situations that were to be compared, in part by administrative considerations of availability of supervisors, noninterference with other projects, etc. Systematic sampling was used extensively within the localities for many of these studies and the samples were appraised critically in the reports. One of the studies was an 11 percent national sampling of workers on relief in March 1935, with proportions varying from state to state, providing very detailed tabulations by age, occupation, and education. The accuracy of the sample results was verified by comparison with subsequent complete tabulations and a table of sampling errors was given [38].

(3) The Study of Consumer Purchases was conducted in 1935-36 by the Bureau of Labor Statistics in 32 cities and the Bureau of Home Economics in 19 small cities and 206 villages or rural counties [39]. The study was a work relief project and other agencies participated in the planning and analysis. In selecting the localities, available data on their economic characteristics were used to achieve a high degree of representativeness; within each locality families were selected systematically from city directories, the schedules of the Real Property Inventory (New York City), the 1934 WPA Census of Chicago, and similar lists. The smaller areas were canvassed completely. From the first simple schedules obtained for 783,000 families a second sample was selected for interviewing about details of income and expenditures. This sample was designed to provide adequate numbers of families in certain relatively rare categories without taking more families in the more prevalent categories than were needed for comparisons between the various types and kinds. The sampling procedure was complicated by the necessity of interviewing in such a manner that if the work were terminated at any stage in the process families that had been visited would still be a representative sample. This project far exceeded any previous study of its kind in size and complexity. The procedures and problems have been described rather fully in the reports and elsewhere. They follow a basic plan that was prepared by the Social Science Research Council in 1929 [40].

(4) The National Health Survey was conducted by the U.S. Public Health Service in 1935-36. It covered 83 cities in 18 states with sampling ratios varying from 1 in 38 for New York City to complete canvass of the smaller places. The sampling was done by dividing each city into small areas of less than 1,000 persons each using the 1930 Census enumerators' maps as the basis, and then drawing a systematic sample of these small districts for complete canvass. The Survey supplied a great amount of data on illness and medical care. Its schedules were also tabulated in great detail by the Social Security Board to obtain

cross classified information about family composition and characteristics for actuarial studies.

(5) The Michigan Census of Population and Unemployment was taken early in 1935 by the State Emergency Welfare Relief Commission as a WPA work relief project [41]. The sampling procedure involved complete enumeration of all cities between 3,000 and 40,000 inhabitants and a 20 percent random sample of smaller places and rural townships. Cities of more than 40,000 inhabitants were sampled by taking all dwellings with certain predesignated house numbers according to a scheme developed by S. A. Stouffer. The reports not only provided detailed data on unemployment in various parts of the state, but also unusual information on migration that was subsequently analyzed by John N. Webb and Albert Westefeld [42].

(6) The Minnesota Income Study was a work relief project sponsored by the Minnesota Resources Commission and operated in 1938 and 1939 using a complete system of stratified sampling with a systematic selection of two-square-mile areas in rural districts [43].

(7) The New York City Youth Survey was made by the Welfare Council of New York City in 1935 taking every hundredth household from the lists of the 1934 Real Property Inventory supplemented by a similar sample of dwellings built after the Inventory [44]. It was also a WPA work relief project. All members of these households between 16 and 24 years of age were interviewed about their leisure activities, employment, schooling, and other personal data.

(8) The Continuous Work History Sample is a four percent sample of workers covered by old age insurance maintained by the Social Security Board in connection with its records of earnings and contributions to the insurance scheme [45]. Beginning with 1937 all reported earnings and related information are compiled for this group of workers. They were selected by use of certain combinations of digits in their account numbers. The sample provides annually distributions by age, sex, race, and taxable wages that are needed for actuarial purposes as well as useful for general economic studies.

(9) The 1937 Enumerative Check Census of Unemployment was conducted by the Director of the National Unemployment Census who had been ordered by an act of Congress to undertake a voluntary census of the unemployed by mail [46]. The Check Census was designed to determine what proportion of the unemployed had actually registered and thereby provide more accurate estimates of the volume of unemployment than had been available previously. The sample was a systematic selection of all households in two percent of the postal carrier routes in the U.S.A. About 18 percent of the population, not served by carrier delivery, was excluded but was taken into account in the estimates.

The 510,000 households in the sample were visited by postal carriers to obtain data on employment and unemployment and a special count was made of the number of workers and unemployed persons for the routes that constituted the sample. From these data it was determined that 71 percent of the unemployed had registered and estimates were prepared for separate classes of workers with corresponding calculations of the sampling errors of the estimates. The analysis of the estimates and their errors was especially noteworthy.

These examples have brought us up to the point at which we approach developments that will be discussed by other speakers on this program.

The foregoing examples exhibit progress in utilizing sampling technique for statistical surveys of exceedingly important national problems. They reflect the trend toward greater use of detailed statistical data and research as a basis for determining policies and establishing programs on a city and state wide basis as well as nationally. They are impressive for the size of the samples taken, but there were also many smaller samples of comparable importance. In two ways, they were a result of the depression: (a) in the urgent demand for statistical data generated by the emergency and by the New Deal programs, and (b) in the funds and relief personnel that were made available. However, these influences tended toward haste and inadequate preparation, toward the use of less efficient sampling schemes in order to meet problems of supervision and the restrictions on expenditures for travel by interviewers and toward types of inquiries that were suitable for the kinds of interviewers that were available. Problems of cost and economical operation were considered but were affected by the primary purpose of providing useful work for unemployed white collar workers. This tended to discourage intensive analysis or research studies not of a relatively routine character since the amount of employment provided in relation to the kind and amount of supervision required was smaller for such project. Finally sampling problems received only secondary emphasis because the other problems connected with the survey were much more serious and time consuming. The results of these surveys were of very great value and to a large extent this was due to the use of sampling procedures.

Coupled with the direct effect of depression problems in stimulating the use of sampling was the trend of thinking among statisticians and social scientists who, during the late 20's and early 30's, became increasingly interested in applying error theory to the analysis of time series and other data [47]. This trend of interest was well represented among many of the men who participated in the work of the Committee on Government Statistics, and recommended in 1933-35 greater and better use of sampling procedures in government statistical work for such purposes as (1) checking the coverage, accuracy, and completeness of censuses, (2) testing the representativeness of current reporting systems (crops, cost of living, payrolls, employment, production, etc.), (3) supplementing the regular schedules in censuses to relieve them of overloading, (4) obtaining prompt national and regional figures on such subjects as unemployment, and (5) making very detailed intensive studies in which the emphasis is on the analysis of relationships [48]. The Committee also recommended that the staff of the Central Statistical Board and the Bureaus of Agricultural Economics, Census, and Labor Statistics include expert advisers on sampling procedure whose formal professional training in the technical phases of sampling was comparable to that of chemists, physicists, or biologists in the government service. These agencies did move to accomplish what had been recommended. They experimented with trial surveys of unemployment, agriculture, construction, retail prices, and basic studies of the problems involved in developing practical sampling systems [49].

By this time sampling practice had outgrown the simpler methods of random and systematic sampling and was developing complex sampling systems specially designed to fit the nature of the population that was being sampled, the costs and administrative factors, and the principles of efficient design that had evolved in scientific experimentation and industrial production.

V. EXPERIMENTAL COMPARISON OF METHODS AND DEVELOPMENT OF COMPLEX SAMPLING PROCEDURES

Modern sampling practice utilizes a variety of devices and methods which have been tested experimentally and incorporated into sampling theory. The aim of modern practice is to select those methods which, when combined in the most appropriate manner, will constitute a sampling system that is economical, convenient, and accurate in the situation for which it was designed. Among the devices available for the construction of such a sampling system are: (a) the technique of selecting items at random, (b) the determination of the kind of sampling unit that will have the most desirable properties, (c) subdivision or stratification of the population in an advantageous manner, (d) the use of variable sampling proportions, (e) subsampling and multistage sampling in which the sample is itself sampled in turn, (f) drawing two or more units from each ultimate subdivision to permit estimation of the sampling error from their differences, (g) use of information provided by variables that are correlated with the one that is being studied, and many other procedures.

It is difficult to find the original source of many of these techniques but it is clear that they were, to a large degree, crystallizations of notions suggested by common sense, customary practices, and practical problems encountered in sampling. Discovery of their importance and success in combining them into efficient sampling schemes has required a clear understanding of their properties and a considerable amount of ingenuity. Beyond that, careful experimentation has been necessary to test their relative merits under various conditions. Their development has occurred primarily in connection with agricultural experimentation; mass production industry; and large scale surveys.

(*a*) *Agricultural experimentation:* One of the first steps in the experimental comparison of sampling schemes was the study of heterogeneity in the fertility of a field and the consequent correlation of adjacent portions. In 1910 and 1911 very important work was reporced by "Student," Wood and Stratton, and Mercer and Hall, [50], leading to the conclusions that in field trials there is always some experimental error, that it can be reduced by taking a large number of small plots and that, in the comparison of two varieties or treatments, it can be reduced by taking these plots in pairs, adjacent to each other, and using one plot out of each pair for each variety or treatment. The latter device was adopted so as to take advantage of the correlation between adjacent areas. Plots of different sizes and shapes were compared using the results of a careful harvesting and measurement of mangolds and data on 400 reports of duplicate plots reported by other investigators. These results led to similar experiments in England, America, and elsewhere [51]. Various systematic sampling schemes were proposed and tried experimentally, both for locating plots and subsampling the plots in order to reduce the labor of harvesting them completely. In 1929, Clapham and Wishart published the results of experimental comparisons between certain random and systematic methods of sampling a field of potatoes or cereals and analyzed the factors that contributed to the error of the estimate of yield [52]. In the same year Smith and Prentice published results of a field study in which they took samples of soil and then subsampled for laboratory examination. In their analysis of the results they estimated the increase in error due to subsampling [53]. R. A. Fisher, whose advice and methods of statistical analysis had been followed in many of the experiments in the 20's and later, developed complex systems of experimentation of great efficiency and revolutionized experimentation [54]. Yates and Zacopanay published in 1935 a thoroughgoing analysis of the efficiency of sampling a field including comparative labor costs [55]. A number of manuals were published expounding the principles and practice of field experimentation. This work of the agronomists and statisticians had direct applications to large scale surveys.

(b) *Mass production industry:* Another source of complex sampling procedures was the development of inspection and quality control in mass production. In 1923, engineers at the Western Electric Company, manufacturers of telephone equipment, began to apply probability theory to the inspection of telephone exchange equipment. In the refinement of manufacturing processes and progressive adoption of more rigorous specifications, the cost of inspection had become very burdensome and sampling procedures offered possibilities of reducing the amount of inspection without serious loss of control over the quality of the product. Different sampling schemes were compared on the basis of their probability of acceptance curves and tolerances were established for the proportion of defective pieces in a lot. Thereafter sampling procedures were developed in more complex forms to achieve a reduction in the amount of inspection to the minimum necessary under the conditions of manufacture and the specifications to be met [56]. Technical problems more or less peculiar to the telephone industry made it the pioneer in the applications of mathematical statistics to industrial processes, including not only inspection but quality control and probability analysis of operating problems under Shewhart, Dodge, Molina, Fry, and others [57].

(c) *Communication of new techniques:* These developments in agriculture and engineering had both direct and indirect effects on sampling survey practice. They provided principles of design and contributed to the growth of applied mathematical statistics. Still there were many practical problems and obstacles that delayed the immediate extension of the methods developed for field trials and manufacturing to large scale surveys. One of the obstacles was the relative lack of communication between statisticians engaged in different types of work. An important step toward facilitating communication was taken by the Royal Statistical Society in 1933 when it formed the Industrial and Agricultural Research Section and published the proceedings of meetings of the Section as a *Supplement* to the Society's *Journal.* The Section brought together mathematical statisticians, engineers, agriculturists, and others interested in practical applications of newly developing statistical theory and stimulated greatly interest in sampling. E. S. Pearson presented a notable paper on "Problems of Industrial Sampling" at a meeting of the Section early in 1934 [58]. A few months later, Jerzy Neyman read before the Society a comprehensive theoretical and critical paper on stratified sampling and purposive selection. He cited among several examples, a survey of the working population in Poland for which he had designed a stratified random sample of districts [59]. A number of other papers by various authors, published from time to time in the Supplement, included discussions of sampling practice, In this period interest in sampling practice was heightened in Great Britain by Shewhart's visit in 1932 (a factor in the formation of the Industrial and Agricultural Research Section) and in America by the visits of Fisher, Neyman, Yates, and Cochran (1936-38) as well as by American students who studied in England.

(*d*) *Large-scale surveys:* Another root of complex sampling practice is to be found in the progressive improvement of large scale surveys of crop yields and acreages, the labor force, and other economic and social facts. A pioneering survey in India in 1923 by J. A. Hubback went almost unnoticed until Mahalanobis discovered and republished it in 1946 [60]. Hubback undertook to estimate the yield of rice by objective measurements in the fields instead of by the methods based on the judgment of farmers and brokers. He used as sampling units very small areas measured off by a wooden frame in every field his agents found being harvested as they walked a fixed route from each selected center. Hubback was acquainted with Student's and Bowley's sampling work. His report emphasized the advantage of random selection of many very small areas as economical and necessary to obtain a valid estimate of error.

In the period 1928-31, J. O. Irwin examined the British crop estimating system and made recommendations for improving the sampling methods [61]. Subsequent work was done by Cochran. In the United States, W. F. Callander, J. B. Shepard, C. F. Sarle and others began to plan in the late 20's for a sample of farms to be used for various studies of acreage, production, livestock, etc. Sarle made comprehensive appraisals of the methods used to obtain farm-price data in 1927 and crop-yield estimates in 1932 and proposed a program of studies aimed at testing and improving the various sampling operations then in use [62]. Several studies of this type were made by state agricultural statisticians.

In July 1936, the Bureau of Agricultural Economics and Iowa State College held a conference on sampling attended by many of the state agricultural statisticians as well as staff from Washington and Ames [63]. R. A. Fisher who was lecturing at Ames participated in the discussions. The conference set forth a great variety of practical problems and outlined specific projects for research to help solve them. Some of these projects were started at Ames in cooperation with the Bureau, some were undertaken in Washington. The program has been continued up to the present time with additions and modifications. It has resulted in a remarkable series of reports and developments of methodology [64] as well as the establishment of a sampling staff within the Iowa State College Statistical Laboratory at Ames under G. W. Snedecor, A. J. King, and R. J. Jessen.

During the same period, P. C. Mahalanobis and his associates at the Indian Statistical Institute in Calcutta developed large scale sampling methods to a high degree of technical perfection and utilized them for surveys of jute acreage, agricultural finance, famine conditions, radio listening, and many other subjects. The theoretical system associated with these practical applications has been set forth fully by Mahalanobis in his publications and lectures [65]. Mahalanobis has made careful experimental measurements of the human errors in sampling surveys by use of "interpenetrating samples" and other means.

From 1936 on the Research Division of the U.S. Bureau of the Census under C. L. Dedrick made extensive studies of problems in population sampling, utilizing the 1930 Census returns. The results have not been published. In 1939, Stephan, Deming, and Hansen designed a sampling scheme that was used in the 1940 Population Census to draw a 5 percent sample totalling 7 million people. This method was based on experimental tabulations of 1930 Census data and on a variation of systematic sampling [66]. The sample was used to obtain preliminary estimates of regular census statistics as well as tabulations of the special questions that were asked only of the people in the sample. Subsequent complete tabulations verified the accuracy of the estimates.

During the same period experiments were conducted in a number of cities and towns by the Research Division of the Work Projects Administration, under H. B. Myers and J. N. Webb, to test out methods of conducting sample survey of unemployment. On the basis of these trials a monthly nationwide survey was started in December 1939 utilizing a subsampling procedure in approximately 60 counties. Its scope was enlarged during the war to include housing and labor questions and in 1942 it was transferred to the Census Bureau where it was developed further as the Monthly Survey of the Labor Force [67].

With the Monthly Survey and other sampling surveys required by the war agencies, the Census Bureau had a sizeable program of sampling work and developed, under Morris H. Hansen, a staff of sampling specialists who designed sample schemes, and studied their performance in practice. Data from the 1940 Census for very small areas were made available for use in sampling and Sanborn maps showing every street and building in a large part of the cities in the United States were obtained [68]. In addition airplane maps were obtained for a large part of the United States and procedures were developed for employing them in sampling.

A further step was taken in the development of sampling facilities when the Bureau of Agricultural Economics established at Iowa State College in 1934, a project to design and prepare a Master Sample of farms for use in its surveys [69]. This Master Sample is a stratified selection of very small areas based on airplane photographs and highway maps. The Census Bureau joined in its development and had it extended to include other than farm areas.

This most recent developments in complex sampling procedures have been described very inadequately and many other recent developments have been omitted for want of space. Other speakers on the program will present the current stage of theory and examples of present day practice.

VI. DEVELOPMENT OF USE OF JUDGMENT SELECTION AND CONTROLS IN SAMPLING

All attempts to obtain data by sampling involve some choices based on the purposes of the study, knowledge of the material, and experience with the methods of observing and recording similar data in previous studies. For example Bowley's choice of Reading for his survey in 1912 and Clapham's choice of certain fields at Rothamsted for his experiments on sampling design were not random selections. Yet the conclusions drawn from such samples as these are commonly generalized to apply to other periods of time and other places. Some investigators have depended entirely on their knowledge and judgment to make the final selection of the items that constitute the sample, others have insisted that judgment should only be used to subdivide the population into strata and units of sampling, the actual choice of the sample being made at random. These views have been presented in various forms and the differences have not yet been fully resolved. Research is in progress to determine some of the facts and develop the principles that are involved in appraising the performance of various sampling procedures based on these conflicting views.

Jensen, in his report in 1926, described a number of surveys made by "purposive" sampling in which districts were selected which, taken together, matched the population in certain characteristics known from a previous census. The theory of purposive selection was developed by Bowley in 1926, and discussed by Jensen in 1928 [70]. Neyman compared purposive and stratified random sampling in 1934 giving special attention to experience reported by Gini and Galvani. Mangus described a variety of purposive sampling of counties used in studies of the FERA in 1934 [71].

The development of public opinion polling and market research involved many interesting problems of social psychology and survey procedure that fall outside the scope of this paper, but it also led to the use of many different sampling schemes prominent among which were those termed "quota control" sampling. The earliest work was done with samples chosen to meet the convenience and very limited resources of the research personnel. In this respect they were like most other surveys conducted at the time. Robinson has reviewed in detail the history of straw votes and opinion polls to 1932 [72]. He also analysed the *Literary Digest* poll which, in spite of its great size, was badly biased by its method of mail balloting. Several analysts were able to improve the *Literary Digest* results by relatively simple adjustments, [73] but, in field surveys in which interviewers are left free to use their judgment in selecting a representative group of respondents, a better procedure appeared to be to regulate the selection of the sample by assigning to each interviewer a quota of persons of various kinds so as to obtain the correct proportions of men and women, old and young, rich and poor, etc., in each city or district that was canvassed. This procedure was developed by Cherington, Roper, Gallup, Crossley, whose sampling surveys of opinion became widely known after the election of 1936, and by market research and opinion survey organizations, whose studies received less publicity [74]. The general principles of these and related methods have been discussed by Brown, Franzen and Robinson, by Cassady, and in two symposia [75]. Quota methods have been employed in the surveys of the National Opinion Research Center [76], the Office of War Information and the British Wartime Social Survey [77]. Most of the opinion polls and market research studies around the world at present employ quota methods or related procedures of sampling. These methods have been adopted partly by imitation of the more widely known surveys and partly because they are attractive with respect to costs and practical convenience.

VII. General View of the Present Period

Current uses of sampling will be described by other speakers at this session. For the most part these uses were connected with the war which proved to be an even greater accelerating stimulus than the depression of the 30's. They included surveys of employment, housing, prices, shortages of consumer goods, dealer inventories, public reaction to wartime measures, characteristics of selectees, the attitudes of soldiers, effects of bombing in England, Germany and Japan, public opinion in many countries, inspection of election registers in Greece, surveys of saving and spending, employee attitudes, reading and radio listening, social and psychological factors affecting fertility in the U.S. and in Britain, etc. In manufacturing and procurement, sampling methods were perfected for quality control and inspection of materials, parts, and finished products. Market research and opinion studies by private research agencies and business concerns developed rapidly, with considerable emphasis on measuring public relations and employee attitudes. The volume of work increased and many continuing or repetitive surveys were established, making it possible to spread the costs and hence devote greater resources to the design of each sample than had been possible for infrequent sampling surveys.

Other speakers will describe the more recent developments that grew out of these earlier accomplishments. The *Journal of the Royal Statistical Society* and its *Supplement*, the *Journal of the American Statistical Association*, *Sankhyā*, the *Annals of Mathematical Statistics* and many of the scientific and professional publications in various fields have published the increasingly numerous articles that reflect the progress of applied sampling. The reader should review them, beginning in 1940 or earlier, and search articles that report particular studies as well as those that are principally devoted to discussions of method.

Great as these advances have been, they are only the forerunners of a broad expansion in the use of sampling in the future. Further improvements in method are also to be expected. Much experimental work is being done both within and outside government bureaus. The Committee on the Measurement of Opinion, Attitudes and Consumer Wants, appointed jointly by the National Research Council and the Social Science Research Council, is making a comprehensive study of sampling methods as they are used in its field. The United Nations Statistical Commission has established a Sub-Commission on Sampling which will contribute greatly to the development of sampling surveys in many countries. The International Statistical Institute can perform a great function in extending the use of sampling procedures in the many statistical activities throughout which it exerts its leadership and influence.

Notes and References

The following abbreviations will be used:

J. Agr. Sci.	*Journal of Agricultural Science*
J. A. S. A.	*Journal of the American Statistical Association*
J. R. S. S.	*Journal of the Royal Statistical Society*
Phil. Trans.	*Philosophical Transactions of the Royal Society of London*

[1] Even in the comprehensive *Handbook of Mathematical Statistics* (Houghton Mifflin, 1924) although considerable space was given to time series none was given to the procedures of drawing samples. As an example of the neglect

of sampling in one among many fields of science, see: Quinn McNemar's critical review, "Sampling in Psychological Research," *Psychological Bulletin*, 37 (1940), pp. 331-365.

[2] Edmund Halley, "An Estimate of the Degrees of the Mortality of Mankind Drawn from Curious Tables of the Births and Funerals at the City of Breslau," *Phil. Trans.* (1693), pp. 596-610.

[3] J. B. Lawes and J. H. Gilbert, " On the Home Produce, Imports, Consumption, and Price of Wheat," *J. R. S. S.*, 43 (1880), pp. 313*f*.

[4] E. C. K. Gonner, "The Population of England in the Eighteenth Century," *J. R. S. S.*, 76 (1912), pp. 261-296. See pp. 279-280.

[5] Henry S. Shryock, Jr., "Methods of Estimating Postcensal Populations, " *Am. Jour. Public Health*, 28 (1938), pp. 1042-1047.

[6] Moheau, "*Recherches et considerations sur la population de la France, 1778*" (republished by Libraire Paul Geuthner, Paris, 1912). Paul E. Vincent, "French Demography in the Eighteenth Century," *Population Studies*, 1, (1947), pp. 44-71. Harald Westergaard, *Contributions to the History of Statistics*, London, P. S. King & Son, 1932, pp.79-99. John Koren, *History of Statistics*, N.Y., Macmillan, 1918, pp. 257-261.

[7] *The Crop and Livestock Reporting Service of the United States*, U.S. Department of Agriculture, Miscellaneous Publication 171, (1933). John Koren, *op. cit.*, pp. 642-652.

[8] It can, perhaps, be traced best by reviewing the successive volumes of the publications of the International Statistical Institute and the various national statistical societies. See also John Koren, *op. cit.*

[9] Faith M. Williams and Carle C. Zimmerman, *Studies of Family Living in the United States and Other Countries*, U.S. Department of Agriculture, Miscellaneous Bulletin 224 (1935).

[10] Leah H. Feder, *Unemployment Relief in Periods of Depression*, New York, Russell Sage Foundation, 1936, p. 38.

[11] *Ibid.*, pp. 77-78.

[12] *Ibid.*, pp. 81-83.

[13] *Ibid.*, p. 228.

[14] *Unemployment in the United States*, U.S. Bureau of Labor Statistics, Bulletins 172 (1915) and 195 (1916).

[15] Frederick E. Croxton, *Unemployment in Columbus, Ohio, 1921 to 1925*, U.S. Bureau of Labor Statistics, Bulletin 409, Employment and Unemployment Series, 1926.

[16] J. Frederic Dewhurst and Ernest A. Tupper, *Social and Economic Character of Unemployment in Philadelphia, April, 1929*, U.S. Bureau of Labor Statistics, Bulletin 520 (1930).

[17] *Unemployment in Buffalo, November 1929, November 1930; November, 1931*, New York State Department of Labor, Special Bulletins 163, 167, 172 (1930–32)

[18] *Unemployment in Syracuse, November, 1931,* New York State Department of Labor, Special Bulletin 173 (1932).

[19] *Hagerstown Morbidity Studies,* U.S. Public Health Service, Public Health Reports, 41 (1926), pp. 2069-2088, (Reprint 1113).

[20] Committee on the Costs of Medical Care, Report 25-6: Incidence of Illness ...Among Family Groups.

[21] Selwyn D. Collins and G. St. J. Perrott, "The Economic Depression and Sickness," *J. A. S. A., Proceedings,* 29 (1934), pp. 47-51.

[22] Claude E. Robinson, *Straw Votes,* New York, Columbia University Press, 1932, pp. 47*f*.

[23] *E.g.,* Stuart A. Rice, *Quantitative Methods in Politics,* New York, Knopf, 1928. W. F. Willcox, "An Attempt to Measure Public Opinion about Repealing the Eighteenth Amendment," *J. A. S. A.,* 26 (1931), pp. 243-261.

[24] A. L. Bowley, "Address to the Economic Science and Statistics Section, British Association for the Advancement of Science," *J. R. S. S.,* 69 (1906), pp. 540-557. "Student," "The Probable Error of a Mean," *Biometrika,* 6 (1908), pp. 1-25. L. H. C. Tippett, *Random Sampling Numbers,* Tracts for Computers No. XV, Cambridge University Press, 1927.

[25] E. Keller, *The Principles and Practice of Sampling Metallic Metallurgical Materials,* U.S. Bureau of Mines, Bulletin 122, 1916. James G. Osborne, "Sampling Errors of Systematic and Random Surveys of Cover-type Areas," *J. A. S. A.,* 37 (1942), pp. 256-264. F. X. Schumaker, "Statistical Method in Forestry," *Biometrics Bulletin,* American Statistical Association, 1 (1945), pp. 29.

[26] Adolph Jensen, "The Representative Method in Practice," *Bulletin* of the International Statistical Institute, 22 (1928), pt. 1, pp. 386*f*.

[27] W. B. Bailey, *Modern Social Conditions,* 1906. F. S. Chapin, *Field Work and Social Surveys,* New York Century, 1920, p. 116.

[28] Adolph Jensen, *op. cit.,* pp. 394-422.

[29] A. L. Bowley, "Working-class Households in Reading," *J. R. S. S.,* 76 (1913), pp. 672-701. A. L. Bowley and A. R. Burnett-Hurst, *Livelihood and Poverty,* London, Bell & Sons, 1915.

[30] A. L. Bowley, "The Application of Sampling to Economic and Sociological Problems," *J. A. S. A.,* 31 (1936), p. 474. D. Caradog Jones, "Evolution of the Social Survey in England since Booth," *American Journal of Sociology,* 46 (1941), pp. 818-826. D. Caradog Jones, "The Social Survey of Merseyside," *J. R. S. S.,* 94 (1931), pp. 218-250. P. Ford, *Work and Wealth in a Modern Port,* London, George Allen and Unwin, 1934. London School of Economics, *New Survey of London Life and Labour,* London, P. S. King; 9 Vol., 1930-35. A. L. Bowley, "New London Survey," *J. R. S. S.,* 92 (1929), pp. 530-547.

[31] John Hilton, "Enquiry by Sample, an Experiment and its Results," *J. R. S. S.,* 87 (1924), p. 544. John Hilton, "Some Further Enquiries by Sample," *J. R. S. S.,* 91 (1928), pp. 519-530.

[32] *Bulletin* of the International Statistical Institute, 25 (1931), pt. 2, p. 121.

[33] B. F. George, "A Sample Investigation of the 1931 Population Census," *J. R. S.* S., 99 (1936), p. 147. Day Monroe, *Chicago Families*, University of Chicago Press, 1932.

[34] Frank A. Ross, "On Generalization from Limited Data," *Social Forces*, 10 (1931), pp. 32-37. Samuel A. Stouffer, Sociology and Sampling," in L. L. Bernard (ed.), *Fields and Methods of Sociology*, N.Y., Ray Long and R. R. Smith, 1934. Samuel A. Stouffer, "Statistical Induction in Rural Social Research," *Social Forces*, 13 (1935), pp. 505-515. Frederick F. Stephan, "Practical Problems of Sampling Procedure," *American Sociological Review*, 1 (1936), pp. 569-580. Frederick F. Stephan, "Representative Sampling in Large-scale Surveys," *J. A. S. A.*, 34 (1939), pp. 343-352.

[35] Margaret H. Hogg, "Sources of Incomparability and Error in Employment-Unemployment Surveys," *J. A. S. A.*, 25 (1930), pp. 284-294.

Margaret H. Hogg, *The Incidence of Work Shortage*, N.Y., Russell Sage Foundation 1932.

[36] M. B. Givens, "An Experiment in Advisory Service," *J. A. S. A.*, 29 (1934), p. 394. *Government Statistics*, Social Science Research Council, Bulletin 26, (1937), p. 132.

[37] David L. Wickens, *Financial Survey of Urban Housing*, U.S. Bureau of Foreign and Domestic Commerce, 1937.

[38] For example T. C. McCormick, *Comparative Study of Rural Relief and non-Relief Households*, Work Progress Administration, Division of Social Research, Washington, 1935. P. M. Hauser, *Workers on Relief in the United States, March 1935*, Work Progress Administration, 1939.

[39] Bureau of Labor Statistics; Study of Consumer Purchases, Bulletins 637 to 649 and 691. Bureau of Home Economics, Family Income and Expenditures, U.S. Dept. Agr. Misc. Pub. 339, 345, 356, 370, etc. Washington, D.C., 1939.

[40] The Special Committee to Formulate a Plan for a Cost of Living Study, 1929, sought to get an adequate sample of each income class and family type to permit comparisons and regression analysis.

[41] *Michigan Census of Population and Unemployment*, State Emergency Welfare Relief Commission, Lansing, 1936.

[42] J. N. Webb and Albert Westefeld, "Labor Mobility and Relief," *Monthly Labor Review*, January 1939, Reprint R866.

[43] See Conference on Research in Income and Wealth, (National Bureau of Economic Research 1943), *Studies in Income and Wealth*, Volume 5, part 2, chapter 10.

[44] Nettie Pauline McGill and Ellen Nathalie Matthews, *The Youth of New York City*, N.Y. Macmillan, 1940.

[45] Conference on Research in Income and Wealth, *op. cit.* ch. 8.

[46] Calvert L. Dedrick and Morris H. Hansen, Enumerative Check Census, Vol. IV of the Census of Partial Employment, Unemployment, and Occupations, Final Report on Total and Partial Unemployment, 1937, Washington, 1938.

[47] Examples are: H. R. Tolley, "Economic Data from the Sampling Point of View" *J. A. S. A., Supplement,* 24 (1929), pp. 69-72. Holbrook Working and Harold Hotelling, "Applications of the Theory of Error to the Interpretation of Trends," *J. A. S. A. Supplement,* 24 (1929), pp. 73-89. Bruce D. Mudgett, "The Application of the Theory of Sampling to Successive Observations not Independent of Each Other," *J. A. S. A. Supplement,* 24 (1929), pp. 108-117. Jacob Perlman, "Some Problems in the Construction of an Index Number of Rents," *J. A. S. A. Supplement* 29 (1934), pp. 25-32. Hugh S. Hanna, "Adequacy of the Sample in Budgetary Studies, *J. A. S. A. Supplement,* 29 (1934), pp. 131-134.

[48] Committee on Government Statistics and Information Services, *Government Statistics,* Social Science Research Council Bulletin 26, 1937, pp. 51, 82, 85, 95.

[49] First Annual Report of the Central Statistical Board, 1935, pp. 24-25.

Third Annual Report of the Central Statistical Board, 1937, p. 14.

C. L. Dedrick, "Progress of the Census," *J. A. S. A.*, 31 (1936), p. 737.

U. S. Bureau of Agricultural Economics, Proceedings of Conference on Statistical Methods of Sampling Agricultural Data, July 14-17, 1936, Iowa State College, Ames, Iowa. (Mimeographed).

[50] "Student," "The Distribution of the Means of Samples which are not Drawn at Random," *Biometrika,* 7 (1909-10), p. 210. T. B. Wood and F. J. M. Stratton, "The Interpretation of Experimental Results," *J. Agr. Sci,* 3 (1910), pp. 415-440. W. B. Mercer and A. D. Hall, "The Experimental Error of Field Trials," *J. Agr., Sci.*, 4 (1911), pp. 107-132.

[51] For Example, T. Lyttleton Lyon, "Some Experiments to Estimate Errors in Field Plot Tests," *Proceedings, American Society of Agronomy,* 3 (1911), pp. 89-114. J. W. Day, "The Relation of Size, Shape and Number of Replications of Plots to Probable Error in Field Experimentation," *J. Amer. Soc. Agron.*, 12 (1920), pp. 100-105. T. Eden, "The Experimental Errors of Field Experiments with Tea," *J. Agr. Sci.*, 21 (1931), p. 546; and the bibliography in *J. Amer. Soc. Agron.*, 22 (1930), 1056.

[52] A. R. Clapham, "The Estimation of Yield in Cereal Crops by Sampling Methods" *J. Agr. Sci.*, 19 (1929), 214-235. J. Wishart and A. R. Clapham, "A Study in Sampling Technique," *J. Agr. Sci.*, 19 (1929), pp. 600-618.

[53] A. M. Smith and E. G. Prentice, "Investigation on *Heterodera schachtii* in Lancashire and Cheshire," *Annals of Applied Biology,* 16 (1929), pp. 324-339.

[54] R. A. Fisher "The Arrangement of Field Experiments," *Journal of the Ministry of Agriculture,* 33 (1926), pp. 503-513. R. A. Fisher, *The Design of Experiments,* London, Oliver and Boyd, 1935.

[55] F. Yates and I. Zacopanay, "The Estimation of the Efficiency of Sampling, with Special Reference to Sampling for Yield in Cereal Experiments, *J. Agr. Sci.*, 25 (1935), pp. 545-577.

[56] Harold F. Dodge and Harry G. Romig, *Sampling Inspection Tables,* N.Y.. John Wiley & Sons, 1944, pp. 1-5.

[57] Walter A. Shewhart, *Economic Control of Quality of Manufactured Product*, New York, D. Van Nostrand Co., 1931. T. C. Fry, *Probability and its Engineering Uses*, D. Van Nostrand Co., 1928.

[58] E. S. Pearson, "Problems of Industrial Sampling, *Supplement to the J. R. S. S.*, 1 (1934), 107-151.

[59] Jerzy Neyman, "On the Two Different Aspects of the Representative Method," *J. R. S. S.*, 97 (1934), 558-625.

[60] J. A. Hubback, "Sampling for Rice Yield in Bihar and Orissa," *Sankhyā*, 7 (1945-6), 281—(reprinted from Bulletin 166, Agr. Res. Inst., Pusa, 1927).

[61] J. O. Irwin, "Crop Estimation and its Relation to Agricultural Meteorology, *Supplement to J. R. S. S.*, 5 (1938), pp. 1-45.

[62] C. F. Sarle, *Reliability and Adequacy of Farm-price Data*, U.S. Department of Agriculture Bulletin 1480, 1927. C. F. Sarle, *Adequacy and Reliability of Crop-yield Estimates*, U. S. Department of Agriculture Technical Bulletin 311 (1932).

[63] Proceedings of Conference on Statistical Methods of Sampling Agricultural Data, U.S. Department of Agriculture, 1936 (Mimeographed).

[64] Irvin Holmes, *Research in Sampling Farm Census Methodology, Part I*, U.S. Department of Agriculture, Agricultural Marketing Service, 1939.

Raymond J. Jessen, "An Experiment in the Design of Agricultural Surveys," *Journal of Farm Economics*, 21 (1939), pp. 856-863.

Arnold J. King and Glenn D. Simpson, "New Developments in Agricultural Sampling," *J. Farm Econ.*, 22 (1940), pp. 341-349.

C. F. Sarle, "The Possibilities and Limitations of Objective Sampling in Strengthening Agricultural Statistics," *Econometrica*, 8 (1940), pp. 45-62. Arnold J. King and Emil H. Jebe, *An Experiment in Preharvest Sampling of Wheat Fields*, Iowa State College, Research Bulletin 273, 1940. Arnold J. King, Dale E. McCarty, and Miles McPeak, *An Objective Method of Sampling Wheat Fields to Estimate Production and Quality of Wheat*, U.S. Dept. of Agr. Technical Bulletin 814 (1942).

Raymond J. Jessen, *Statistical Investigation of a Sample Survey for Obtaining Farm Facts*, Iowa State College Research Bulletin 304 (1942).

Raymond J. Jessen and Earl E. Houseman, *Statistical Investigations of Farm Sample Surveys Taken in Iowa, Florida and California*, Iowa State College Bulletin 329 (1944).

[65] Consult the reports in *Sankhyā*, including the pertinent parts of the annual reports of the Indian Statistical Institute, and P. C. Mahalanobis, "On Large-Scale Sample Surveys," *Phil. Trans.*, Series B, 231 (1944), pp. 329-451.

[66] Frederick F. Stephan, W. Edwards Deming, and Morris H. Hansen, "The Sampling Procedure of the 1940 Population Census," *J. A. S. A.*, 35 (1940), pp. 615-630.

[67] John N. Webb, M. Starr Northrup, and Stanley L. Payne, "Practical Applications of Theoretical Sampling Methods," *J. A. S. A.*, 38 (1943), pp. 69-77.

Lester R. Frankel and J. Stevens Stock, " On the Sample Survey of Unemployment," *J. A. S. A.* 37 (1942), pp. 77-80.

[68] A. Ross Eckler and E. P. Staudt, "Marketing and Sampling Uses of Population and Housing Data," *J. A. S. A.*, 38 (1943), pp. 87-92. Morris H. Hansen and W. Edwards Deming, "On Some Census Aids to Sampling," *J. A. S. A.*, 38 (1943), pp. 353-357. A. N. Watson, "Use of Small Area Census Data in Marketing Analysis," *Journal of Marketing*, 6 (1942), pp. 42-47.

[69] A. J. King and R. J. Jessen, "The Master Sample of Agriculture," *J. A. S. A.*, 40 (1945), pp. 38-57.

[70] Adolf Jensen, "Purposive Selection," *J. R. S. S.*, 91 (1928), pp. 541-547.

[71] A. R. Mangus, "Sampling in the Field of Rural Relief," *J. A. S. A.*, 29 (1934), pp. 410-415.

[72] Claude Robinson, *Straw Votes*, Columbia University Press, 1932.
Claude Robinson, "Recent Developments in the Straw-poll Field," *Public Opinion Quarterly*, 1 (1937), No. 3, pp. 45-56, and No. 4, pp. 42-52.

[73] For example, L. S. Crum, "On Analytical Interpretation of Straw Vote Samples," *J. A. S. A.*, 26 (1931), pp. 243-261.

[74] Daniel Starch, "Factors in the Reliability of Samples," *J. A. S. A.*, *Supplement*, 27 (1932), pp. 190-201. American Institute of Public Opinion, *The New Science of Public Opinion Measurement*, New York, (pamphlet). The Fortune Survey, *Fortune*, quarterly. Elmo Roper, "Sampling Public Opinion," *J. A. S. A.*, 35 (1940), pp. 325-334.

Archibald M. Crossley, "Straw Votes in 1936," *Public Opinion Quarterly*, 1 (1937), pp. 24-35. D. Katz and H. Cantril, "Public Opinion Polls," *Sociometry*, 1 (1937), pp. 155-179.

[75] Hadley Cantril (ed.), *Gauging Public Opinion*, Princeton University Press, 1944. Theodore H. Brown, Raymond Franzen, and D. E. Robinson. "Design, Size and Validation for Marketing Research," *Journal of Marketing*, 10 (1946), 221-233. Albert B. Blankenship (ed.), *How to Conduct Consumer and Opinion Research*, N.Y., Harper, 1946. Ralph Cassady, Jr., "Statistical Sampling Techniques and Marketing Research," *Journal of Marketing*, 9 (1945), pp. 317-341.

[76] *Interviewing for NORC*, National Opinion Research Center, Denver, 1946.

[77] H. Box and G. Thomas, "The Wartime Social Survey," *J. R. S. S.*, 107 (1944), p. 151.

Résumé

Les premiers exemples d'enquêtes par sondage ont été décrits dans des aperçus historiques par *Jensen*, *Westergaard* et autres auteurs. La conception d'une méthode "représentative" distincte d'un recensement complet, d'une part, et des études partielles au moyen d'"enquêtes" ou de "monographies", d'autre part, a été clairement mise en lumière

au cours de la discussion d'une communication de *Kiaer* à la session de Berne de l'Institut International de Statistique en 1895. Les opinions furent extrêmement partagées et, après discussions aux trois sessions suivantes, l'Institut donna en 1903, sous réserve, son approbation à cette méthode.

Bowley recommanda vivement les enquêtes par sondage dans un discours en 1906, et utilisa le sondage systématique dans son étude du paupérisme dans quatre villes anglaises (1912-1913), et dans d'autres études. Des techniques de sondage systématique furent employées au Japon (1923), aux Etats-Unis (1927), et en Angleterre (1935) pour certains dépouillements de bulletins de recensement. En Angleterre, des enquêtes nationales sur le chômage furent menés par sondage systématique des fiches des Bureaux de Placement (1921-1928). Cette méthode fut utilisée aux Etats-Unis (1931, 1934, 1935, 1937) dans des enquêtes sur le chômage, les budgets familiaux, les problèmes de la jeunesse et, sous une forme spéciale, dans le recensement démographique et agricole de 1940. La théorie du sondage systématique a été traitée par *Madow* (1944), *Cochran* (1946) et par d'autres auteurs. La méthode a été largement employé dans le sondage de minerals, de charbon et autres matières premières, dans le contrôle de la qualité de produits manufacturés et, comme un des éléments de procédés de sondage plus complexes, dans de nombreuses enquêtes agricoles et économiques.

Bien qu'ayant été très efficaces, les méthodes systématiques n'ont pas donné entière satisfaction pour certains buts. C'est pourquoi des méthodes d'échantillonnage plus complexes furent élaborées, en particulier dans le domaine de la recherche agronomique. Le problème du choix de méthodes de sondage qui permettent de tenir compte des difficultés pratiques, du coût et de l'exactitude, fut étudié à la fois expérimentalement et théoriquement; des améliorations dans les méthodes utilisées furent dévelopées notamment dans des articles de *Hubback*, *Fisher*, *Clapham*, *Wishart*, *Yates*, *Cochran* (1923), *Bowley* (1926), *Neyman* (1934), *Mahalanobis* (1937), *Sarle*, *King*, *Jessen*, et leurs collaborateurs (1938), *Stephan*, *Deming*, *Hansen*, et *Hurwitz* (1940), et de nombreux autres auteurs. Aux Etats-Unis, ces progrès sont dus, en partie, à une amélioration des données de base, telles que spécifications géographiques très détaillées des données de recensements, photographies aériennes, l'amélioration des cartes routières et plans de la propriété foncière mis à jour pour les nouvelles constructions. Les études du marché, d'opinion publique et certaines études sociologiques ont suivi un développement quelque peu différent basé sur une méthode généralement connue sous le nom de "contrôle par quota" utilisée notamment par *Roper*, *Gallup* et *Crossley* dans les études d'opinion publique qui depuis 1935 ont donné lieu à un intérêt général pour les sondages.

L'emploi des sondages s'est beaucoup développé pendant la dépression de 1930 et pendant la guerre, les périodes de crise créant un besoin urgent de statistiques en de nombreux domaines et sous des formes nouvelles que la statistique traditionnelle n'est pas à même de procurer en temps utile.

SAMPLING OF HUMAN POPULATIONS

by Morris H. Hansen
Statistical Assistant to the Director
Bureau of the Census (United States)

This paper is not intended to present any new results; its purpose is to review some methods of designing a sample that will yield estimates of the numbers that would result from a complete census of the population under consideration. Particular attention will be given to methods for obtaining a sample from a city or other local area. For example, the problem might be to ascertain how many persons reside in an area, or perhaps what proportions of them have specified ages, employment status, income, attitudes, or other characteristics. The procedures involved in collecting the information from a particular individual are assumed to be the same whether the coverage is complete or on a sample basis. These procedures should be such that accurate and useful results would be obtained from a complete census using these procedures. The results of such a complete census will be referred to as the "true" values of the quantities involved.

The reliability of the results obtained from a sample survey depends not only on the size of the sample, but also and with equal importance on the sample design—that is, on the way in which the sample is selected and the way in which estimates are prepared from the sample survey returns. Good sample design results from the effective use of available resources. Possible resources include not only such facilities as staff, maps, and tabulating equipment, but also statistical information and other knowledge of the population to be sampled, together with existing sampling theory. While the methods to be employed will depend on the particular information desired, the fundamental sampling principles are the same for problems involving the estimation of characteristics of individuals, families, farms, business establishments, or other populations.

It will be assumed that an acceptable sample design must satisfy the following requirements.

(1) The first requirement is that the design must be such as to provide practical certainty that the estimate made from the sample does not differ by more than a prescribed amount from the value to be estimated, assuming that the procedures specified by the design are substantially followed in the actual selection of the sample and in the computation of the estimate. In other words, the design used should be

such that, in addition to making estimates of the desired characteristics from the sample, one can also derive from the sample itself or from prior knowledge measures of the precision of those estimates. A measure of precision will indicate the maximum error that may reasonably be expected. This requirement is fulfilled if the selection of the sample is carried out in such a way that the probability of inclusion in the sample is known for each element in the population (farm, individual, or whatever is being sampled).

(2) The procedures prescribed must be so simple and straightforward that they can be carried out substantially as specified within the time limits allowed and with the resources available. When such methods are used we can achieve in practice results that agree with the sampling theory. Theory and practice are thus brought together.

(3) From among the alternative designs which fulfill the two preceding requirements, one should choose that design which will produce results of maximum reliability per unit of cost. Most of the recent advances in the theory of sampling human populations and institutions have been directed at the accomplishment of this last requirement. It is this requirement that calls for locating and making effective use of available resources. Theory is not now available to guide one uniquely to the best method, but the available theory is exceedingly useful in guiding one to the choice of an effective method.

It should be emphasized that for methods that fulfill the first two requirements mentioned above, the reliability of the results can be objectively evaluated, and therefore does not depend on assumptions or judgment as to the representativeness of the sample. Samples that meet the above requirements will reflect a population as it is, taking account of changes that may have taken place in a population subsequent to a census; and, moreover, methods that meet these requirements can be applied to produce results of any desired precision by taking a large enough sample even where no census of the population under consideration ever has been taken.

Methods that do not meet these requirements and that are essentially expressions of judgment may be available at very low cost. Thus, suppose we want for the city of Washington an estimate of the proportion of dwellings that are owner-occupied. One way of estimating this proportion is to ask someone whom we regard as well-informed on housing conditions to give us an estimate based on his judgment. This approach might produce very accurate results, but experience indicates that it often does not. Even if the expert's judgment is good it is likely that others to whom we may present the results will not share our confidence in the soundness of a judgment unsupported by adequate objective evidence.

Instead of relying on the judgment of a single individual we might consult several such persons, and perhaps have them suggest representative neighborhoods of the city for us to investigate, or we might purposively choose "typical" families for interview. These methods still reflect the judgments of the persons choosing the sample. We might go further, and insist that the sample be selected in such a way that certain classes of families be represented in the sample in proportion to their estimated numbers in the population being sampled, but with the final choice of the particular individuals to be included left to the judgment or convenience of the interviewers subject only to the restriction that he must meet his quotas by classes of families. By thus increasing the number of sources of information, however, we have not essentially changed the logical basis for our confidence in the final estimate. The sample estimate is still dependent on the judgments exercised in the final selection, and we cannot evaluate objectively the soundness of those judgments.

All of these procedures are rejected in the light of the requirements indicated above. The estimates resulting from these procedures lack the assurance given by methods for which objective measures of precision can be derived. There is an advantage in having the more objective information that is provided by random sampling methods, for the objective measures of precision that result mean that one can often avoid making erroneous decisions or choosing a wrong and perhaps costly course of action. Moreover, the statistical theory applicable to such methods not only provides measures of precision, but also guides one in the choice of efficient sample designs.

Sampling theory has long been available for what we might consider the ideal circumstances for taking a sample. Most of you are probably familiar with the simple formula

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

as the variance of a mean of a random sample, where n is the size of sample used and σ^2 is the variance of the original population from which the sample was drawn. The variance of a sample estimate provides a measure of the precision of that estimate. Thus, one can be almost certain that with a moderately large random sample the mean of the sample does not differ from the true mean being estimated by more than three times its standard deviation (the square root of its variance) computed according to the formula above. Not only does the above formula provide a measure of the reliability of the results of a random sample, but it also provides a basis for determining in advance of drawing the sample approximately how large the sample should be in order to attain the precision required.

The importance of the word "random" in this connection must not be overlooked. A sample is not random simply because the person selecting the sample is not conscious of any bias. Randomness must be deliberately introduced into the selection process by use of a table of random numbers or by some equivalent device.

The simple formula given above, with a slight modification to take into account the finiteness of the population, will apply to sampling the population of a city if an unrestricted random sample of *n* persons is drawn from a list of the inhabitants of the city. If such a list were always available, and if considerations of cost did not enter into the problem of collecting information from the members of the sample, there would be no need for further sampling theory. For most practical purposes, a sample drawn from such a list would yield a smaller sampling error than a sample of the same size drawn by any other reasonable alternative methods.

The need for investigating alternative methods of sampling arises from the fact that such lists usually are not available and are too costly to prepare. Moreover, even were such resources available, there is no assurance that information can be secured from the entire sample by a reasonably inexpensive procedure. The reasonable alternative methods usually require a larger sample than the "ideal" procedure indicated above, if the same reliability is to be assured. But ordinarily the additional cost introduced by the increased size of sample will be small compared to the excessive costs and time required if the approach used involved the preparation of a complete and up-to-date list of all persons in the city and the interviewing of a sample drawn at random from such a list.

Whenever a random sampling method is used a formula for the variance of the sample estimate can be constructed. The formula will depend on the sampling procedure followed. Such formulas always involve the size of the sample, although not usually as simply as in the formula quoted above. Consequently, variance formulas provide a basis for determining the sample size required to achieve a specified precision and for evaluating the precision from the sample. Moreover, if unit costs for the various operations involved in taking the survey are known, the unit costs together with the formula for the sampling variance provide a basis for choosing the optimum sampling design—that is, the design which gives the greatest precision for a given cost.

While the preceding discussion may seem to point to the use of very complex procedures, yet the actual selection of a random sample can be relatively simple. A random sample can be drawn by any of a number of expedients, even with limited resources with which to work.

Suppose, for example, we want to estimate the proportion of employed persons in a city and other similar characteristics. It is not essential to have a great deal of information or to introduce complicated processes in order to obtain a random sample of persons. We might have only a reasonably up-to-date map of the city which shows the street boundaries. The city map might be a published map, or an aerial photograph, or a sketch of the block boundaries within the city. The only requirement is that the map be sufficiently accurate that individual blocks or other small areas can be identified in such a way that the field workers can locate exactly the areas that are selected.

To obtain a random sample of blocks, one can first number the blocks serially. The blocks may be numbered in any convenient sequence throughout the city. Assume for simplicity that it has been determined that for the design under consideration a sample of 500 blocks will yield an estimate of the proportion of employed persons with the required reliability. Of course, the size of the sample needed will depend on the accuracy required and on the method of sampling.

If there are 5,000 blocks in the city, we can obtain a sample of 500 blocks from among them by including in the sample one block in ten. One way of obtaining the random sample of 500 blocks is to draw 500 numbers between 1 and 5,000 from a published table of random numbers and include in the sample the blocks having these serial numbers. A census of the selected 500 blocks would then be taken. The census would yield for the sample blocks both the total number of persons and the total number of employed persons, along with other desired information. An estimate of the proportion employed could then be obtained by computing the ratio of these two numbers from the sample.

The above very simple procedure will yield results of measurable accuracy, since in this particular instance each individual in the population has a known chance of being included in the sample. However, this procedure will often not be the best one for dealing with the problem. There are methods of modifying the sampling procedure which will ordinarily yield the required information from the sample at considerably less cost.

A simple and inexpensive modification of the procedure for drawing the sample would be to introduce a certain amount of stratification in connection with the selection of the blocks. One obvious mode of stratification that can be readily imposed is geographical. Often there are fairly wide differences in some of the characteristics of persons living in different sections of the city and a geographical stratification will reduce the effect of these differences on the sampling error.

Such stratification is easily introduced if the blocks are numbered systematically, section by section throughout the city, instead of in an arbitrary order. With the blocks so numbered stratification can be accomplished by employing a systematic random sample. This is accomplished if we choose a random number between one and ten, include the block with that number in the sample, and every tenth one thereafter. Thus, if the random number chosen were 2, the blocks with serial numbers 2, 12, 22, etc., to 4992 would be included in the sample. The estimating procedure will be the same as that indicated earlier; and the sampling error for this procedure usually will be somewhat smaller for the same size of sample.

A second and perhaps more important use of stratification is to identify by inquiry, in advance of drawing the sample, those blocks that have very large aggregate populations, say, blocks having more than approximately 50 dwelling units. It is particularly important to identify large blocks with unusual characteristics such as those with large hotels, lodging houses, hospitals, jails, military establishments, etc. The large blocks can then be sampled separately from the other blocks of the city. The work of identifying these special types of blocks can be facilitated by discussion with local groups such as city officials, the fire department, the police, postal employees, and other informed local groups. One can thus make use of an available resource in the form of the existing knowledge of various people in the community.

It is not necessary to obtain an accurate identification of this special group of blocks in accordance with some rigid definition of "large" blocks. If we fail to include certain of these blocks in the special group, the sample of the remaining blocks will still reflect their contribution to the total, but the reliability of the sample result will be lower than if they were all correctly classified beforehand.

Other criteria of stratification might be used in addition to the two that have been discussed. The investigator should recognize, however, that extensive additional stratification beyond introducing some of the fairly obvious criteria is not likely to produce a substantial increase in precision in the estimates of general population characteristics such as the proportion employed, or distributions by age, sex, and other classes. Moreover, any of a number of different stratification procedures may be equally effective, so that it is usually not worthwhile to debate extensively the comparative merits of competing methods of stratification.

A method for making more significant gains in the efficiency of the sample is to reduce the number of interviews per block and to increase the number of blocks in the sample. The possible increase in precision can be seen by considering as an illustration two ways of obtaining a sample from a small section of a city that contains ten blocks and with an average of 25 families per blocks. First, consider a sample drawn from this area by preparing a list of all families in the area and choosing 25 families at random from the list. Such a sample would be scattered over the entire area. It seems plausible that such a sample of 25 families will yield a more representative cross-section of the area than would a sample of 25 families obtained by drawing a single block at random and including all families within the selected block. The single block will yield a less reliable sample because neighboring families tend to resemble one another more than will families throughout the area. It appears reasonable, and it can be shown that when this is true the families from a single block will not as adequately represent the various classes of families in the area under consideration as would a more widely scattered sample of the same size.

The block sample approach was introduced because, as was suggested earlier, we do not have a list of families from which to draw a random sample, and because it may be much too expensive to prepare such a list. However, the reasoning just used suggests that we could obtain a better estimate for a given size sample by scattering the sample more widely. This could be accomplished by taking two half-blocks instead of one full block, or better yet by taking smaller segments of blocks, or by otherwise increasing the number of areas in the sample and decreasing the average size. Some statistical theory is available to guide one in determining the optimum size of segment to be used.[1] The gains by this line of approach may be very substantial, but often with the resources available, it is not feasible to designate a sample consisting of segments of blocks.

There are a number of ways other than subdividing the blocks which make it possible to have a larger number of blocks represented in the sample without increasing the number of families to be interviewed. One way is to select a larger sample of blocks but then to introduce a second stage of sampling within the selected blocks. This second stage of sampling is commonly referred to as subsampling. In a subsampling procedure one may list all the dwellings or other units to be subsampled in the sample of blocks, and select for interviewing a sample of the units listed. Then, if a sample of 10 percent of the persons is desired one might select every fifth block and then interview the persons living at every other address or dwelling within the selected blocks. One might then say that the effective size of the block is no longer the original block but the size of the sample drawn from the block. As was indicated earlier, a 10 percent sample drawn in this manner gives considerably more reliable results than would a 10 percent sample of blocks completely enumerated.

---

[1] See references 6, 7, 8, 15, and 18.

When such a subsampling procedure is used one must keep the method of listing the units within the sample blocks as simple and economical as possible. In determining the kind of subsampling units to be listed one need only to remember that they should be readily identifiable, that they should be small, and that every person in the block must live in one of the listed units. For example, it is satisfactory to list dwellings and to interview all persons living at a sample of these dwellings provided that the dwellings listed are not too large (say, do not contain except in rare instances more than 40 persons). Apartments within structures containing (or appearing to contain), say, 5 or more apartments should be separately listed. Within institutions and large lodging houses the unit of listing may be either a person or a room, whichever is easier to identify and list. If rooms are listed the occupants of the rooms selected for the sample will be covered.

The "best" or optimum determination of the proportion of blocks to be drawn into the sample from the various strata and the subsampling ratios to be used in the selected blocks is given by the available sampling theory.[2] The optimum ratios will depend on the relative costs of various operations, and on the amount of variability between blocks. If there is a relatively large variation between blocks in the characteristics being estimated, then, as one might expect intuitively, the theory points to the use of a larger proportion of blocks and a smaller subsampling ratio than would result if the variation between blocks is relatively small. This last statement has an important implication for sampling from the stratum of large blocks that has been discussed. These blocks not only are large in size but also the variation between them ordinarily will be relatively great. It follows that the proportion of those blocks included in the sample should be larger than the proportion included from the remaining blocks.

While statistical theory may be used to determine the optimum block sampling ratios and subsampling ratios, it is not necessary to adhere rigidly to such optimum values. One may depart moderately from the optimum values without significantly increasing the cost or the sampling error. Thus, the investigator has considerable latitude in selecting block sampling ratios and subsampling ratios so as to simplify the operations and processing of the sample. One such simplification, for example, is to impose the condition that the results of the sample be self-weighting; that is, the estimate of the proportion of employed persons be simply the ratio of the number of employed persons in the sample to the total number of persons in the sample. This is accomplished by making the product of the block sampling ratio and the subsampling ratio the same for the special group of large blocks as for the other blocks.

---

[2] See references 1 and 11.

The fact that it is not necessary to adhere rigidly to the optimum allocation in order to achieve most or a substantial part of the gains that may result from the application of the optimum theory means that some simple rules for drawing a block sample for a city can be given that have fairly wide applicability. Thus, an approximation to the optimum sampling ratios will often be accomplished by merely

(a) using a uniform over-all sampling ratio in all strata;

(b) determining the block sampling ratio and the subsampling ratio in the stratum of "small-blocks" so that on the average about five families will be included per sample block; and

(c) making the proportion of blocks to be drawn from the "large-block" stratum equal to

$$\sqrt{\frac{\text{average number of families per block in the "large-block" stratum}}{\text{average number of families per block in the "small-block" stratum}}}$$

multiplied by the proportion of blocks to be drawn from the "small-block" stratum.

Assume, for example, that a sample of 5 percent of the households of the city is to be taken, and that the average number of families per block in the "small-block" stratum is approximately 25, and in the "large-block" stratum is approximately 100. To obtain an average of about five families per block from the "small-block" stratum the subsampling ratio would be 5/25 = 20 percent. Since the percent of dwelling units to be included in the sample is to be the same in both strata and equal to 5 percent, the proportion of blocks to be drawn from the "small-block" group is $\frac{5 \text{ percent}}{20 \text{ percent}} = 25$ percent. The proportion of blocks to be drawn from the "large-block" stratum would then be

$$25\sqrt{\frac{100}{25}} = 50 \text{ percent,}$$

if we follow the rule just given; and the subsampling-ratio for the "large-block" stratum would be

$$\frac{\cdot 05}{\cdot 50} = 10 \text{ percent.}$$

The introduction of subsampling may make it possible to attain the accuracy with perhaps a 3 percent sample of families that would result from a 10 percent sample if the blocks were to be completely enumerated. Such a decrease in sample size for equivalent reliability is not unusual when one uses a reasonably close approximation to the optimum design.

It should be emphasized that the preceding discussion of methods of sampling from a city has been concerned with the problem of estimating proportions, percentage distributions, or other average characteristics from the sample and not with the problem of estimating the total number of persons in the population, or the total number of employed persons, or other similar aggregates. The principal difficulty in estimating *totals* from block samples as described above arises from the sensitivity of such estimates to wide variations in the number of families per block. Estimated ratios are not affected so severely by the variation in size of block, because both the numerator and the denominator of the ratio often tend to vary together so that their respective sampling errors are likely to offset one another partially in the computation of the ratio.

If blocks were relatively uniform in size the procedures described above would yield good estimates of totals. But because blocks ordinarily vary widely in size, some further resources must be found and used to reduce the effect of variation in size of block, or a further price must be paid in the form of modifying the sampling procedure or of substantially increasing the size of sample if estimates of totals are likely to be made with a reasonable degree of reliability.

The additional resources that may prove particularly helpful are counts of the approximate number of dwellings in each block in the city. These may be available from a relatively recent census, or may be obtainable from aerial photographs or from other maps or records. If such approximate measures of size cannot be obtained from available records, it may be worthwhile to obtain them by means of a field survey procedure in which each block in the area is visited and quick count of the approximate number of dwellings in the block recorded. Such an operation can be carried out with satisfactory accuracy at a fraction of the cost that would be involved in preparing an accurate listing of the addresses of all the dwellings in an area. Once measures of size are obtained for the blocks, the blocks may be stratified by size. Sampling the blocks in each stratum and then subsampling dwellings, as described earlier, will result in a sample that will yield much more reliable totals, for a given size of sample, than would result from the sample design suggested earlier in which only the very large blocks were identified in the stratification.

This procedure was followed a few years ago in the Bureau of the Census to obtain highly reliable estimates of population totals for a series of metropolitan areas that had been subject to tremendous growth during the war and for which the prior census figures had become so out-of-date as to be of little use for current planning. These surveys, using fairly large samples, gave estimates of the total population that could confidently be expected to be in error by less than 2 percent for each of the metropolitan areas.[3]

There are methods other than stratifying by size of block for reducing the effect of variability in size of block on estimated totals, that are discussed elsewhere.[4]

Another type of resource may sometimes be available that can be used to achieve the desired results at lower cost. Sometimes a relatively recent city directory or other moderately complete listing of the dwellings in an area is available. Such listings may be used effectively and without biasing the sample results, even though they are not complete, provided they do not contain duplicate listings for the same address. This type of resource can be used by proceeding in the following way:

The dwellings in the city can be regarded as divided into two major strata for sampling purposes: (a) The dwellings that are listed in the directory; and (b) The dwellings that are not listed in the directory. The sample from the stratum consisting of those listed in the directory can be sampled in accordance with the ideal procedure mentioned earlier of sampling from a list. In some instances they may be sampled in small clusters from the list if evaluation of cost considerations, together with the increased sampling variance resulting, point to such an approach. In such a sample some addresses drawn may no longer represent dwellings. This is the case if a building has been converted to non-residential use, or torn down or destroyed, or has otherwise disappeared. Such addresses should be omitted from the sample. No substitution should be made for these since the sample of addresses should reflect such disappearances.

The sample of the remaining dwellings in the city—those that do not appear in the directory—should be drawn by a block sampling procedure similar to the one outlined above. The procedure for this stratum of dwellings not in the directory will be to check the directory for each sample block and pick up from this sample block only those dwellings that do not appear in the directory. If there are in the city some blocks that have an especially large amount of new housing, it may be desirable to define these blocks as constituting another special stratum.

It should be pointed out that a design that is adequate for estimating population aggregates is also an effective design for estimating proportions. Thus, the introduction of controls on size of block will have the effect of increasing somewhat the accuracy of estimates of proportions, although the importance of such controls here is not as great as it is in estimating totals. For estimating totals careful control on the size of blocks may enable us to make such estimates with an accuracy equal to that obtainable from a sample that is several times as large but used without such control on the size of blocks.

---

[3] See reference 1.

[4] See for example, references 6, 8, 11, and 24.

It should be remarked finally, with respect to the problem of sampling from a city that although designs can be used for which the actual sampling operations can be carried out in a simple and straightforward manner, the determination of the necessary size of sample, and the reliability of results from a sample are technical problems calling for careful study and application of available sampling theory, much of which has not yet appeared in published form. If efficient and reliable samples are to be obtained, this work must be put in the hands of persons trained in sampling theory and practice.

This discussion has centered on the problem of making estimates for local areas such as cities or counties. When estimates are desired for the country as a whole, or for other broad areas, many other considerations must be taken into account. Factors which are relatively unimportant in sample design for local areas may now become extremely important, and conversely. For example, in nation-wide surveys the convenient, or the optimum primary sampling unit to be used may often contain as many as 100,000 or more persons, while in local area surveys primary units with as few as 200 persons may be considered large. Stratification plays a much more significant role in designs for national than for local surveys, although even for national surveys its importance has often been exaggerated. Travel costs, mapping costs, the use of past information, all have a more prominent part in the design of national samples. The type of sampling used in such surveys is affected by the organization of the field operations, particularly whether the administration of the work is centralized or decentralized. It is impossible within the limitations of the space of this paper to elaborate on these principles but they are discussed in the available literature.[5]

Similarly, the sampling of farms, business establishments, or certain types of data for individuals or families such as average or aggregate income, where a few individuals or units contribute a considerable part of the total, call for other special sampling procedures that have not been treated in this paper. While the fundamental sampling theory to be applied is the same in designing samples for these purposes, the points for special emphasis are sometimes quite different. In these problems, for example, stratification, and the use of special lists, assume an especially important and significant role. Mail survey methods are especially important in sampling business establishments, and can be supplemented by other methods in order to yield results of known precision. Some further discussion of these problems and methods is available elsewhere.[6]

---

[5] See especially references 6, 8, 11, 18, and 19.

References

(Some selected articles on or illustrating applications of sampling methods, with particular references to the sampling work of the Bureau of the Census.)

1. *A Chapter in Population Sampling*, by the Sampling Staff, Bureau of the Census, 1947. (For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D. C. Price $1.00.)
2. W. G. Cochran, "Sampling Theory When the Sampling Units Are of Unequal Sizes," *Journal of the American Statistical Association*, Vol. 37 (1942), pp. 199-212.
3. W. G. Cochran, "The Use of Analysis of Variance in Enumeration by Sampling," *Journal of the American Statistical Association*, Vol. 35 (1939), pp. 492-510.
4. W. Edwards Deming, "On Errors in Surveys," *American Sociological Review*, Vol. 9, No. 4 (1944), pp. 359-369.
5. Morris H. Hansen and W. Edwards Deming, "On Some Census Aids to Sampling," *Journal of the American Statistical Association*, Vol. 38 (1943), pp. 353-357.
6. Morris H. Hansen and William N. Hurwitz, "A New Sample of the Population"—Sampling Principle Introduced in the Monthly Report on the Labor Force, Bureau of the Census, 1944.
7. Morris H. Hansen and William N. Hurwitz, "Relative Efficiencies of Various Sampling Units in Population Inquiries," *Journal of the American Statistical Association*, Vol. 37 (1942), pp. 89-94.
8. Morris H. Hansen and William N. Hurwitz, "On the Theory of Sampling from Finite Populations," *Annals of Mathematical Statistics*, Vol. XIV No. 4 (1943), pp. 333-362.
9. Morris H. Hansen and William N. Hurwitz, "Sampling Methods Applied to Census Work," Bureau of the Census, 1946.
10. Morris H. Hansen and William N. Hurwitz, "The Problem of Non-Response in Sample Surveys," *Journal of the American Statistical Association*, Vol. 41 (1946), pp. 517-529.

---

[6] See references 9, 10, 11, and others.

11. Morris H. Hansen, William N. Hurwitz, and Margaret Gurney, "Problems and Methods of a Sample Survey of Business," *Journal of the American Statistical Association*, Vol. 41 (1946), pp. 173-189.

12. Philip M. Hauser and Morris H. Hansen, "On Sampling in Market Surveys," *Journal of Marketing*, Vol. 9, No. 1 (1944), pp. 26-31.

13. Philip M. Hauser and Morris H. Hansen, "Area Sampling—Some Principles of Sample Design," *Public Opinion Quarterly*, Vol. 9, No. 2 (1945), pp. 183-193.

14. Ernest R. Hilgard and Stanley L. Payne, "Those Not at Home; Riddle for Pollsters," *Public Opinion Quarterly*, Vol. 8, No. 2 (1944), pp. 254-261.

15. R. J. Jessen, *Statistical Investigation of a Sample Survey for Obtaining Farm Facts*, Iowa State College Research Bulletin 304, 1942.

16. A. J. King and R. J. Jessen, "The Master Sample of Agriculture," *Journal of the American Statistical Association*, Vol. 40 (1945), pp. 38-56.

17. William G. Madow and Lillian H. Madow, "On the Theory of Systematic Sampling I," *Annals of Mathematical Statistics*, Vol. 15, No. 1 (1944), pp. 1–24.

18. P. C. Mahalanobis, "Sample Survey of the Acreage Under Jute in Bengal," *Sankhyā*, Vol. 4 (1940), pp. 511-530.

19. J. Neyman, "Contribution to the Theory of Sampling Human Population," *Journal of the American Statistical Association*, Vol. 33 (1938), pp. 101-116.

20. J. Neyman, "On the Two Defferent Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, New Series Vol. 97 (1934), pp. 558-606.

21. Frederick F. Stephan, W. Edwards Deming, and Morris H. Hansen, "The Sampling Procedure of the 1940 Population Census," *Journal of the American Statistical Association*, Vol. 35 (1940), pp. 615-630.

22. Benjamin J. Tepping, William N. Hurwitz, and W. Edwards Deming, "On the Efficiency of Deep Stratification in Block Sampling," *Journal of the American Statistical Association*, Vol. 38 (1943), pp. 93-100.

23. F. Yates, "A Review of Recent Statistical Developments in Sampling; and Sampling Surveys," *Journal of the Royal Statistical Society*, Vol. 109, Part 1 (1946), pp. 12-43.

24. *Sampling Plans for a Census of Haiti*, prepared by the Subcommittee on Sampling Techniques, Bureau of the Census, 1942.

25. *Sample Design for Proposed Survey of Population and Housing*, prepared by the Sampling Staff, Bureau of the Census, 1946.

Résumé

Cet exposé n'a pas pour but de présenter de nouveaux résultats mais plutôt de passer en revue quelques-unes des méthodes de sondage donnant des estimations dont on puisse calculer l'exactitude. Seuls seront

pris en considération ici les plans d'échantillonnage remplissant les conditions suivantes:

1) Il doit être pratiquement certain que la différence entre l'estimation et le paramètre à estimer ne dépasse pas une certaine quantité fixée d'avance.

2) La procédure envisagée doit être simple et directe.

3) Parmi les différents plans remplissant ces deux conditions, les plan choisi devra donner la plus grande exactitude pour un coût donné.

Si les deux premières conditions sont remplies, on peut évaluer objectivement l'exactitude des résultats du sondage, sans dépendre de suppositions ou de jugements sur la plus ou moins bonne représentation que l'échantillon donne de la population. Les plans d'échantillonnage qui remplissent les conditions précédentes doivent utiliser effectivement toutes les sources d'information disponibles, mais la sélection finale de l'échantillon doit etre laissée au hasard. Pour de telles méthodes de sondage il est possible de calculer l'erreur d'échantillonnage, ce qui permet de déterminer le nombre d'éléments que l'échantillon doit comporter afin qu'on obtienne une certaine précision donnée à l'avance. De plus, si le coût unitaire des différentes opérations de l'enquête est connu, il devient possible de choisir la méthode qui donne la plus grande précision pour un coût donné.

Malgré l'apparente complexité de la théorie des sondages, le choix d'échantillons au hasard de différents types, peut en pratique être relativement simple. Une méthode, par exemple, pour obtenir un échantillon au hasard des habitants d'une ville, est de choisir au hasard des "blocs" et de former l'échantillon avec tous les habitants des blocs choisis. Un tel procédé peut, cependant, ne pas être très efficace.

Un échantillon formé de blocs peut, en général, être amélioré en faisant à l'avance une liste de certains blocs ayant une population très élevée et en faisant un sondage particulier de ces blocs. Plusieurs autres procédés de stratification peuvent aussi être utilisés. Des gains significatifs dans l'efficacité de l'échantillon peuvent être obtenus en répartissant l'échantillon plus largement à travers la ville par une augmentation du nombre de blocs dans l'échantillon et par une diminution du nombre d'"interviews" dans chaque bloc. La théorie des sondages permet de déterminer le nombre optimum de blocs à comprendre dans l'échantillon et la proportion de la population de chaque bloc à comprendre dans l'échantillon.

Si l'échantillon composé de blocs doit donner de bonnes estimations de chiffres totaux plutôt que de proportions ou de moyennes, il est important de contrôler les variations du nombre de ménages par bloc. Ceci peut se faire par un compte approximatif du nombre de logements dans chaque bloc de la ville et en classifiant ensuite les blocs suivant ce nombre.

L'importance relative des différents aspects d'un plan d'échantillonnage dépend en grande partie du problème particulier considéré. Par exemple, des facteurs spéciaux doivent être pris en considération quand le problème consiste à choisir un échantillon de fermes, d'établissements commerciaux, ou de ménages, en vue d'estimer des caractéristiques comme le revenu moyen ou le revenu total, cas dans lesquels il est possible que quelques unités constituent une très large part du total.

COMMERCIAL USES OF SAMPLING

by J. Stevens Stock and Joseph R. Hochstim
Opinion Research Corporation
(United States)

I

The purpose of this paper is to report some of the commercial uses of sampling. We shall describe a series of actual examples which show how a commercial research agency meets its sampling problems. As in all sampling problems, the nature of the research objective and the facilities at hand are the all-important conditions which determine the approach.

1. Most commercial agencies must be prepared to tackle an extremely wide variety of sampling problems. No one sampling procedure or principle is applicable in all, or even in a majority, of cases.
2. The actual budget is frequently not commensurate with the actual difficulty of the problem. Often it is necessary to do the best job possible under a very restrictive budget, even though a larger budget would allow for much greater dollar efficiency.
3. Most commercial research problems are geared into current-policy determination. Consequently, research efficiency must sometimes be sacrificed for speed.
4. Although there are important exceptions, commercial research problems are for the most part nonrecurrent. Time series and repetitive studies are conducted, but are relatively small in number.
5. Many research problems must be tackled with limited source material because of the specialized nature of the problems.
6. Many commercial research studies must cover unusual, difficult-to-define universes of people, such as the purchasers of a given commodity, all chemists in the United States, all people who will vote in the next election, etc.
7. Commercial research agencies often find it necessary to study attitudes and opinions which have highly emotional, personal, or moral implications in the respondent's mind.
8. Very often, even in large national studies, the commercial research agency must make detailed analyses for relatively small areas within the country. This makes it necessary to maintain extremely large part-time staffs widely distributed throughout the nation.

Regardless of any pet theories, the sampler has to consider the following:

1. The problems the user of the data needs to have answered.
2. The actions likely to be based on the results of the survey.
3. The degree of precision required.
4. The size of the budget.

In any case, the sample design chosen should provide the user of the data with maximum information at minimum cost.

II

Our first example involves the sampling of purchasers of alcoholic beverages. It is illustrative of the quick, nonrecurrent survey of a difficult-to-define universe on a subject that for some respondents has moral, personal, and political implications. In one particular state a law had been passed increasing the liquor tax. The client, a distilling company, wanted to know who is getting the blame from purchasers of hard liquor for the increase in prices—the state government, the distillers, or the distributors. The answer to this question would determine the direction the client's public relations effort will take and the issues on which it will concentrate.

Early testing of the questionnaire indicated that people were unable or unwilling to give accurate answers as to their purchases of alcoholic beverages, except at the time of purchase. The client, however, was interested in directing his public relations effort at the purchasers of his product so that it would have the greatest impact. It was necessary, therefore, to design the sample so that the number of respondents was approximately proportionate to the amount of sales in the sample areas.

In the state being surveyed packaged alcoholic beverages are sold only in state liquor stores. Data on liquor sales were available for the past year for different districts of the state, for individual towns in each county, and for individual stores. The sample was designed as follows:

The number of interviews with purchasers in the various districts was computed in proportion to the amount of liquor sales in each district. The two largest cities of the state were removed from their districts and treated as separate units. Interviews were allocated to these two cities separately according to the amount of their own liquor sales.

Each district was divided into three groups of counties: those with high per capita consumption, those with medium, and those with low per capita consumption.

Within each of these three groups of counties the towns were stratified by per capita volume of liquor sales. One sample town was selected from each stratum. The number of interviews with liquor purchasers allocated to each sample town was in proportion to the total volume of liquor sales of the county group from which the town was selected. Within each town a list of stores was prepared and arranged in rank order according to volume of liquor sales. Stores were then selected within each sample town systematically with probability of selection proportionate to sales volume.

Interviewers were instructed to stand outside the stores selected and to interview every *n*th person leaving the stores until the allocated number of interviews was complete. In order to assure proper distribution by time of day and day of the week, interviewers were instructed to interview half of their respondents on Friday afternoon and Saturday, and the other half on other days of the week. In addition, the interviewers were instructed to fill half their quotas before 4:00 p.m. each day and half of them after 4:00 p.m. each day.

III

Our next example is typical of a study of the community relations of an industrial firm.

The largest employer in a Midwestern town of about 100,000 population felt that his company had done a good job of public and industrial relations. But in spite of all its efforts the company had a strike. The executives wanted to know very quickly what could be done to correct an apparent hostility of the citizenry of the community. The problem of the survey was to find what the people in the community were thinking about the company and its employees and what the company could do in the opinion of the community to make their plant a better place to work.

The 1940 Census Block Statistics, which were available for this city, were brought up to date in numbers of dwelling units with the aid of the city planner's data on new construction. All blocks were then stratified first by census tract, then by average rent. Within each of these strata they were arranged by the number of dwelling units per block. The sample blocks were selected by systematic random sampling with probability of inclusion proportionate to the number of dwelling units in each block.

Within each block the interviewer was given a starting point and instructed to go in clockwise direction and to interview every *n*th dwelling unit in the assigned block until the quota for that block was filled.

In cases of non-interview because of refusal or because no one was at home, interviewers were told to go to the next dwelling unit, but to continue the count from the dwelling unit originally designated. In order to minimize the bias resulting from the omission of not-at-homes and to give more adequate opportunity for getting employed workers in the sample, interviewers were required to do their interviewing in the evening and over weekends.

Quotas covering the whole city by sex and age were given to the supervisor. By cumulative daily tallies of the interviews returned, the supervisor was able to watch the representativeness of the survey. He assigned varying quotas to the interviewers for the selection of individuals within the sample households when certain age-sex groups in the population appeared to be under-represented according to the over-all quota.[1]

IV

The next example is a survey to measure the attitudes of social science teachers toward basic principles of free enterprise economics. The particularly difficult problem in this case was to get a cross section of a highly specialized population, namely college and high school teachers of government, economics, sociology, and history.

The sample of high school teachers was allotted by states according to data on high school enrollment published by the Office of Education. Interviews were allocated to cities and towns in proportion to the population distribution of urban places within each state. The selection of the 160 cities and towns within states and size-of-town groups was determined by the location of the company's resident interviewers.[2]

---

[1] Recent experience with this combination of areal and quota techniques indicates that much of the interviewer's tendency toward bias is eliminated by the systematic pre-selection of each dwelling unit in which an interview is to take place. The use of quotas for the selection of an individual within each household allows the interviewer very little free choice.

[2] At first glance, it would seem that the selection of towns to suit the convenience of the resident interviewers is quite unrigorous. In this case, however, it is an excellent means of selecting a random sample of towns. Since there is no conceivable correlation between the methods by which part-time resident interviewers are recruited and any attitude characteristic of the social science teachers being studied, we can say that we have selected our sample at random.

It is a common practice in commercial research firms (and others as well, we presume) to base the selection of localities for the cross section on the location of the resident staff. Many suspect this technique of bias. However, as long as the investigator can guarantee that the method of recruiting is in no predictable way connected with the problems studied, we can be assured that the selection of localities is as much at random as if it had been based on the use of random number or some game of chance.

Most commercial research agencies that maintain large part-time resident interviewing staffs recruit them in proportion to the distribution of the population by region, state, and size of town groups within states. In this case the total maximum number of resident interviewers available was approximately 1,000 located in more than 600 places.

As for college teachers, a complete list of all colleges was prepared by type—senior, junior, and teacher colleges—and was arranged by states. The number of teachers was recorded for each college, and the colleges were put into two groups: those with a teaching staff of 200 or more and those with less than 200 teachers. Every *n*th college was selected for the sample.

Since in neither case was there available a total count on any social or economic characteristic of teachers of social science subjects, it was impossible to assign any quotas by characteristics. It was necessary, therefore, to insure a strict randomization of the respondents. Accordingly, interviewers were instructed to obtain a complete list of social science teachers for the sample colleges and all high schools in the towns selected. The interviewers were further asked to alphabetize this list and to interview the first *n* teachers starting with a given letter of the alphabet. Interviewers were given different starting letters.

A separate part of this survey illustrates the use of mail questionnaires in studying a highly specialized universe. A large industrial company had conducted for social science teachers a course of several weeks in some of the social problems faced by business and the techniques used in solving these problems. The purpose of this study was to measure the impact on teachers of impressions obtained in this course.

A complete list of all participants in the course was obtained. Post cards were mailed to participants asking them to state by return mail whether they would be willing to answer a mail questionnaire.

A great majority of the teachers expressed their willingness to answer the mail questionnaires. When the returns were counted, however, slightly less than 50% of all participants actually mailed questionnaires.

Since it was felt that the returned sample might not be representative and that those teachers who had not returned the mail questionnaires might be different in their characteristics and attitudes, it was decided to take a separate sample of non-returners by personal interviews.

The industrial company that gave the course was located in a very large Midwestern city, and about 40% of the teachers attending the course lived within an area of 15 miles from the plant city. It was decided that resident interviewers were to telephone all social science teachers included in the original list who lived within a 15-mile radius of the plant city. They were to inquire whether the teachers had returned the mail questionnaires and if not, to make an appointment for a personal interview.

By this method, interviews with approximately one fourth of the teachers who did not answer the mail questionnaires were secured. It was then assumed that those who did not answer the questionnaires and who lived in the plant city area were representative of all the teachers not answering the questionnaire. Accordingly, the answers obtained by personal interviews were weighted to give proportionate representation to those who had, and those who had not, answered the mail questionnaires.

V

Much commercial research is devoted to studying the style or product preferences of highly specialized groups in the population who are actual or potential customers for a certain commodity. The purpose of such research is to get rough estimates of relative customer preference in order to guide manufacturing and sales policy.

A survey of this kind was recently conducted to determine style preferences for a particular piece of furniture. In order to set up as nearly as possible a realistic buying situation, the study was restricted to three separate groups:

1. Well-to-do respondents were questioned on styles in high-priced items.
2. Average and medium-well-to-do respondents were questioned on medium-priced items.
3. Poor respondents were questioned on styles in low-priced items.

All types of respondents were questioned on a few furniture styles. To simplify the approach and meet the restrictive budget, interviews were conducted only in towns with population of 50,000 or more. All towns having 50,000 or more population were stratified by geographic sections. Interviews were allocated to cities and towns in proportion to their population in groups of 15 or multiples of 15. In the case of small cities, this amounted to arraying them by population size and selecting an ordered random sample of cities. However, instead of selecting every *n*th city, we selected every *n*th person, so that the probability of selecting any one city was determined by the chance of selecting any one person living in the city. Therefore, the probability of selecting each city was proportionate to its population size.[3]

Within each of the cities selected a relatively large primary sample of 100 to 300 blocks was chosen by selecting every *n*th block from the published block statistics. From these large samples of blocks, smaller samples of blocks were selected for assignment.[4]

---

[3] A common practice of commercial research agencies is to assign interviews in groups of, say, 10 or 15, in order to give each interviewer an assignment large enough to make it worth his while to start on the study and small enough so that he can finish his interviewing in the time limits set for the study.

Within each city the primary sample of blocks was divided into three broad groups: high-rent, medium-rent, and low-rent blocks. In the large cities in the sample, where as many as three groups of 15 interviews were assigned, three interviewers were assigned to sample blocks corresponding to each of the three rental groups. In smaller cities, where only one or two sets of 15 interviews were assigned, interviewers were required to interview in sample blocks from (correspondingly) only one or two of the rental types. Thus, no attempt was made in smaller cities to get a cross section of each city since no individual city analyses were required.[5]

Blocks within each of the three broad rental groups in each sample city were selected by stratifying the primary sample of blocks for each broad rental group into three strata according to rent. From each of these three rental substrata, three blocks were selected at random with probability of selection in proportion to the number of dwelling units on each block. Thus, each interviewer was assigned one type of questionnaire in one broad rental group. Further, he was assigned three blocks from each of the three rental substrata (9 blocks in all). The interviewer was instructed to take five interviews from each set of the three blocks in each rental substratum. He was also instructed to go around as many of the three blocks as necessary to contact respondents owning the particular piece of furniture or saying they were in the market for one. If the respondent did not meet this qualification, the interview was discontinued and contact was made with the next dwelling unit. Wherever the respondent owned such a piece of furniture or was in the market for one, the interview was completed.

Various styles of furniture of different price range were judged by respondents (higher-priced pieces by respondents living in the higher-rent blocks, lower-priced pieces by respondents in the lower-rent blocks). On a few pieces of medium price, however, we measured preferences for all three groups for purposes of experimental control.

---

[4] This practice of selecting relatively large primary samples of blocks, then stratifying them by section of the city, rent, and number of dwelling units, for further selection of smaller samples of blocks amounts to *double sampling*. It appears to be a very useful practice where small samples of blocks from relatively large cities are needed. The very practical reason for this kind of double sampling is to save on the burdensome clerical time that would be required to stratify all the thousands and tens of thousands of blocks in a large city. It is necessary only to stratify the primary sample of 100 to 300 blocks.

[5] The reason for this assignment procedure was that a different questionnaire was used for each type of respondent—the well-to-do, the average, and the poor. Because of the complicated nature of the interview and the danger of confusing the questionnaires, only one type of questionnaire was assigned to each interviewer.

VI

A frequent problem for any research agency is the high expense of travel from interview to interview. The next example illustrates a problem in studying farmers' opinions, one of the worst situations from the standpoint of interviewer travel cost.

During wartime, government regulations determined prices and conditions under which much farm produce could be sold by farmers to food manufacturers and processers. There was considerable resentment on the part of the farmers against these regulations. There was also resentment against the manufacturers who in the opinion of some farmers manipulated these regulations to their own advantage and to the farmer's disadvantage.

The purpose of the survey was to study farmers' sentiments on this issue. It was decided to restrict the study to seven counties in a Midwestern state—an area where the client did most of his purchasing.

Interviews were allotted to each county in proportion to number of farms, but in order to save cost, interviewing was limited to every other township.

State highway maps were obtained on which square-mile sections within each sample township were numbered.

The number of sections within each sample township was recorded, and a cumulative total of these sections was obtained. Then the average number of farms per section was computed. In each county the desired number of sections was obtained by dividing the assigned number of interviews by the average number of farms per section. In order to concentrate interviewing still more, two adjacent sections were combined into one double section. The sampling ratio was then obtained by dividing the total number of sections in the sample of townships by the number of double sections wanted. An ordered sample of every *n*th double section was selected. Interviewers were instructed to interview every farmer in the double sections selected, making call-backs wherever the farmer was not found at home.[6]

VII

Frequently firms desire to make opinion studies among their customers, or associations want to poll a cross section of their members. Often in such cases the client firm or association can supply a complete list of the universe to be studied. The sampling problem is then usually very easy; we stratify the names on the list as efficiently as possible and select every *n*th name.

---

[6] This concentration of interviews in double sections is, of course, less efficient in a statistical sense than a sample of single sections or individual farms. In the information-per-dollar sense, however, it is undoubtedly much more efficient. Moreover, it greatly facilitates call-backs.

Shortly before the end of the war, the planning board of a metropolitan area was concerned with the potential postwar employment in the area. During wartime the population of the area had increased very considerably, and it was important to make an estimate of the number of people for whom employment might be provided in the years to come. The budget for the survey allowed interviews with 1,500 employers in the area. A complete set of cards showing each firm, its industry group, and the numbers of its employees was made available by the client. In the area to be studied there were approximately 750 firms having 50 employees or more, which accounted for about 54% of all employees in the region. It was decided to include all these 750 firms in the sample.

Of the remaining 15,000 firms having less than 50 employees every 20th one was selected.

Thus, approximately equal numbers of firms having more and less than 50 employees were selected for the sample. By full coverage of firms that were more important in terms of employment, the statistical error in projecting the average number of employees was reduced considerably.[7]

The cards for employees of 50 persons or less were stratified by detailed industry groups, and were arranged within each group in ascending order from the smallest to the largest firms. Every 20th card of each industry group was then selected, carrying over the count from one industry group to the next.

VIII

The next survey, which uses quota sampling, illustrates a very important and useful technique for studying many different and wide spread populations at the same time.

In this instance the purpose of the survey was to bring out pertinent facts bearing upon the public relations program of a particular industry. A variety of issues was to be covered, and each issue was to be explored thoroughly.

---

[7] This example illustrates the common case in which certain segments of the population to be studied contribute such high variability to the final statistical estimate that it is efficient to sample disproportionately from different strata. This amounts to a very simple application of what is commonly called "Neyman sampling."

The research problem was to describe in detail the attitudes of many different "publics" toward this industry:[8]

1. Manual workers employed in this particular industry.
2. Residents of each of the major areas (each was analyzed separately) in which the industry is concentrated.
3. Separate cross sections of the general population in seven geographical regions.
4. A national cross section of the continental United States.

The national cross section was made up of interviews allotted to approximately 450 localities in proportion to the population of geographic regions, states, and of size of town groups. Each interviewer was assigned quotas by sex, age, race, employment status, and occupation. These quotas were based on 1940 Census data adjusted wherever possible to more recent Government estimates.

In four of the seven geographic regions, the national cross section sample was adequate. In the other three, additional interviews were taken to supplement those obtained in the national cross section. These extra interviews were assigned to the sample localities in proportion to the distribution of the general public in these regions.

Major industrial areas were defined as all counties in which a minimum of 3% of all employed men were engaged in this particular industry. Those counties were stratified according to population and according to precent of men employed in the industry. These substrata were set up so that the aggregate populations of the counties in each were approximately equal. A sample of counties was then drawn at random from each stratum. General-public interviews in the national or regional sample which were taken in counties where more than 3% of the employed men were employed in this industry were included in the sample of major industrial areas. Consequently, the number of interviews allocated to counties in each stratum was reduced by the number of interviews already taken in this stratum in other samples. Since there was considerable overlap between the different areas analyzed, very substantial savings were effected by taking advantage of the duplications. Quotas were assigned by sex, age, race, employment status, and occupation in proportion to the distribution of such characteristics in each major industrial area.

---

[8] More and more in public relations research it has been found useful to study the "concentric circles" of populations toward which the public relations effort is being directed. This puts great premium on developing efficient sampling techniques for overlapping populations. In such cases ingenious and workable methods are needed to take fullest advantage of the additional efficiency gained through making each successive sample contribute to the sample of the next concentric population.

Interviews with industry employees were taken in the sample counties in the major industrial areas in proportion to the employment in the industry in those areas. The remainder of the industry employee sample was allocated among the national sample localities in proportion to the employment in the industry in those localities.[9]

IX

A very interesting and ingenious application of an older technique, which has been largely abandoned in the past ten years, has recently been reported by Benson, Young, and Syze.[10]

This technique, now known as pinpoint sampling, was used by the American Institute of Public Opinion in connection with the 1944 presidential election.

Past election results for five particular states were found to represent a remarkable reflection of national election outcomes for all two-party contests since 1916.

Within each one of these five states, eight precincts were selected for the sample. The precincts within each state were chosen in such a way that, "giving equal weight to each, the average of the percentages of the vote of either major party was virtually the same as the election result for the state." This was found to have been true in three previous statewide elections.

Within the precincts selected interviewers were instructed to interview in every—or every other—dwelling unit, depending on the number of voters in the precincts. The interviewers still retained the freedom of selecting respondents within the family group, except for the limitation of obtaining over-all quotas for sex and age.

Interviews were completed in about one half of the households in the sample election districts. Slightly more than one fourth of the potential voters were interviewed.

---

[9] Note: Quota sampling has been widely condemned in recent years. Principally, its validity can be questioned on the underlying inverse logic on which it is based—namely:

1. A representative cross section will reflect known characteristics accurately;
2. Therefore, a sample which reflects known characteristics accurately is a representative cross section.

In spite of this dubious logic, it can be shown that under certain not uncommon conditions quota sampling may have the highest cost efficiency of any method now known.

[10] Edward G. Benson, Cyrus C. Young, and Clyde A. Syze, "Polling Lessons from the 1944 Election," *The Public Opinion Quarterly*, Winter 1945-46. The authors are indebted to Messrs. Benson, Young, and Syze for their permission to present these findings.

The results show a deviation of 1·9% from the actual national vote. Part of this deviation was due to the selection of states and part to the selection of precincts. Benson, Young, and Syze concluded from the results that improvements could most profitably be made by enlarging the number of sample areas and taking a smaller number of interviews within each area.

We have here a case of sampling by purposive selection.[11]

X

Although the problems already presented show the great variety of situations a commercial research agency has to meet, there are many more entirely different sampling and interviewing situations that cannot be covered in detail here.

Some of the sampling techniques described might vary according to the particular survey situation. In a survey taken in a town and in its surrounding rural area, for neither of which were block statistics available, the sample was designed by drawing routes on a town map and state highway maps. The routes were drawn in such a way as to include different sections in the town as well as the different types of roads in the survey area. Interviewers were to interview in every *n*th house on the right side of the route designated until their quotas were completed. Different starting points were chosen for the routes, some leading from the town towards outlying areas, others starting as far as 15 miles from the town and leading toward it.

Another example of a special situation: Even though interviews are generally done personally with only *one* respondent in any one household, there are cases where it is best to interview husband and wife together in order to duplicate a family buying situation, (for example, when a survey is made on intended purchases of refrigerators).

---

[11] The use of the method of purposive selection has been roundly condemned by many writers on the subject of sampling. Gini and Galvani in an article in *Annali Di Statistica* in 1929 showed empirically that this method was very inefficient. Later, Neyman in his now famous article, "On the Two Different Aspects of the Representative Method," *The Journal of the Royal Statistical Society*, 1934, confirmed theoretically Gini's and Galvani's findings.

Neyman, however, did express the opinion that under certain very rigid conditions purposive selection might be an efficient method of sampling. One of those very rigid conditions was that the technique might "work" when the controls were very highly and linearly correlated with the statistics to be measured. Although Benson, Young, and Syze do not discuss this point, it seems reasonable that the method of pinpoint sampling may have very important possibilities in predicting election results.

Even though most questioning is done by personal interview, there are situations where telephone interviews are satisfactory (for example, a survey with a short and strictly factual questionnaire and a specialized universe all members of which are telephone subscribers).

In other cases, such as in "in-plant studies," self-administered group interviews are used with a sample including every *n*th employee.

In situations where respondents might feel that their statements could place them in some form of jeopardy, it may be necessary to interview on the street to allay any possible suspicion that respondents might be traced.

In cases of a management sample, such as sales managers or public-relations directors of companies with assets of more than $5,000,000, it may be necessary to sample from such sources as *Moody's* and *Standard and Poor*, and from individual state directories, supplemented by whatever information is available through local chambers of commerce.

In a survey with stockholders, where the composition of the universe was unknown, information was obtained on the characteristics of the types of people who are most often owners of stocks by including a question on stock ownership in other large national surveys. The survey with stockholders could then be restricted to these types, with the realization, of course, that the unusual types of stockholders would not be included.

There are numerous other situations—radio surveys, magazine surveys, advertising studies and many others—which require various approaches suited to their special needs.

Looking back at the examples given, we realize that there are many other possible sampling approaches to these problems. The purpose of this paper is to report, not to analyze. These were the approaches that were actually used. They are presented here to illustrate a few of the many commercial uses of sampling.

Résumé

Comme dans tous les problèmes d'échantillonnage, la nature du but de la recherche et les facilités dont on dispose sont les plus importantes conditions qui déterminent la méthode à suivre.

1. Nombre de bureaux commerciaux doivent être préparés à entreprendre l'étude d'une extrêmement grande variété de problèmes

d'échantillonnage. Une seule et même procédure ou un seul et même principe ne peuvent être appliqués à tous les cas, ni même dans la majorité des cas.

2. La difficulté des problèmes posés et les budgets disponibles ne varient pas proportionnellement.

3. La plupart des problèmes sont posés en vue de décisions immédiates de gestion des affaires. En conséquence, la qualité de la recherche doit parfois être sacrifiée à la rapidité d'information.

4. Bien qu'il y ait d'importantes exceptions, les problèmes de recherche commerciale, pour la plus grande partie, ne sont pas périodiques. Des séries ordonnées dans le temps et des études répétées sont relativement peu nombreuses.

5. A part les conditions limitatives de budget et de rapidité d'information, nombre d'enquêtes doivent se fonder sur un matériel de base limité en raison de la nature spécialisée des problèmes.

6. Nombre d'études doivent se rapporter à des populations non usuelles, difficiles à définir.

7. Les bureaux de recherche commerciale se trouvent souvent dans la nécessité d'étudier des attitudes et opinions qui provoquent des réactions émotionnelles, personnelles ou morales parmi les personnes touchées par l'enquête.

8. Très souvent, même dans de grandes études sur l'échelle nationale, le bureau de recherche commerciale doit faire des estimations détaillées concernant des territoires relativement peu étendus du pays. Cela rend nécessaire d'entretenir un très nombreux personnel à horaire partiel réparti sur tout le pays.

Indépendamment de tout point de vue théorique, il faut considérer les points suivants:

1. Quels sont les problèmes auxquels l'usager des données désire avoir une réponse?

2. Quels actes seront probablement basés sur les résultats de l'enquête?

3. Degré de précision requis?

4. Importance du crédit?

En tout cas, le plan choisi doit procurer à l'usager des données un maximum d'informations pour un minimum de frais.

Les auteurs, en une série d'exemples, montrent comment un bureau de recherche commerciale résout ses problèmes d'échantillonnage. Les exemples concernent les types de problèmes suivants:

1. Une étude rapide non récurrente concernant une population difficile à définir, sur une matière causant des réactions d'ordre moral, personnel et politique pour certaines des personnes touchées par l'enquête.

2. Une étude caractéristique des relations externes d'un établissement industriel, utilisant une combinaison des méthodes d'échantillonnage spatial ("area sampling") et par quotas ("quota sampling").

3. Une enquête effectuée à l'aide de questionnaires transmis par la poste, se rapportant à une population très spéciale.

4. Une étude d'un type usuel pour déterminer les préférences de style ou de produits de groupes très spéciaux étant ou pouvant être acheteurs de certains produits, pour servir de guide à la politique de production et de vente.

5. Une étude des attitudes de "cercles concèntriques" de populations à l'égard d'un établissement industriel. Le but de l'étude était de guider, sur nombre de questions de détail, le programme de relations publiques de l'établissement.

6. Une prédiction relative aux élections présidentielles aux Etats-Unis, en 1944, à l'aide de la méthode "pin-point", un type d'échantillonnage à choix raisonné.

Discussion

Professor Mahalanobis:

Professor Cochran mentioned yesterday that sampling statisticians do not read one another's papers. That I think is true, but one reason was the blackout in India during the war. You simply could not get any literature. From India we did send some here, through the U.S. Embassy in Calcutta, up to about 1939 or 1940. Since then, we have been completely out of touch. And this is why I can sincerely say that I think the job done by Professor Stephan is invaluable. I hope it will be followed up, especially coming to more recent times. Going back to the remarks made by Professor Cochran, there has been much work on parallel lines; in this country, in England, in India; but the blackout is still there.

There is one aspect of this parallel development, which is of considerable interest from the point of view of the history of science. In the broad fundamentals, it is encouraging that we have had not only similar but practically identical developments. It must offer a wide basis for confidence that workers experimenting or developing the theory independently in different countries used often not only identical methods, but sometimes even identical terms, like 'grid', which has been used in the United States, as we did in area sampling in India. This is the first postulate of scientific philosophy: If there is a truly objective method, workers independently should come to the same result. We have that great confidence behind us, that in the sampling field, we have had an actual demonstration of the objectivity of the method, because of parallel developments occurring independently in different countries.

Yet, the contrasts are also of considerable interest. One rather sharp difference in the methods developed in the U.S.A. and in India in the same field of area sampling arises from differences in the economic and technological background of the two countries. In the 'master-sample' in the U.S.A., the attempt was made to set up a strong foundation; to prepare a master sample out of which, at any moment, you can draw what you like. In India, we have been far more *ad hoc*, using maps and each year doing practically independent sampling. The reason is that in India we didn't know whether we would be asked to do the work next year, and we didn't have the money to get ready for next year. In this country a vast amount of capital expenditure was incurred in getting the 'master-sample' ready, which saved money in later years. The sample here took considerable time to prepare, whereas in India in one State, we had to organize an agricultural census and census of population in about 23 days. There was no time to make advance preparations like the 'master-sample'. In the paper on commercial uses, we find the same thing illustrated; namely, the need of adapting the theory of the technical method to the particular problem, and even more than that to the resources (in money most often, sometimes also in time) available. But it is good, in a primarily experimental science like sampling, that experimentation should go on along with advance of theory. It does help in clearing up many theoretical points; it does help in offering many new theoretical problems. And it is all to the good, provided the need of unifying and developing fundamental theory is not lost sight of.

Referring to the question of theory, I shall shoot a few questions to my colleagues here. How about stratification? How do you stratify? We should like to know. Some work is being done on that in India. For example: You have a fairly good knowledge of the universe through a previous study. Knowing this, how do you stratify? We always do it by some sort of—well, a truly statistical method—trial and error.

That leads to another type of problem. Stratification obviously depends on the purpose in view. If we are trying to make an estimate of the total production of a crop, or the total population over a whole geographical area, that is one problem. There is another type of problem which I have called, for lack of a better name, a mapping survey, where you do not want merely the total, but the actual distribution of something over space. A simple example would be a land map, where it is not the total acreage under corn or wheat you require but the geographical distribution of the fertility, or the rate of yield per acre. You may have to allocate more observations on the turning points of the yield surface rather than on portions where variation is small. A certain amount of work done in India shows that the stratification may have to be on an entirely different basis. So stratification is not something which is automatically given; it is an intrinsic part of the design of sample surveys. This is a very important theoretical problem.

Professor Stephan has referred to the method of what we call interpenetrating samples. We attach great importance to this method as an effective control. It is simply this: In a great deal of our work with the area method, we split our total sample into two (or sometimes a larger number of) portions. Suppose you are taking, 1,000 farms or 5,000 little areas, for investigation; we arrange this sample in two interpenetrating groups which we shall mark in blue and yellow. You collect information for the blue group and the yellow group independently. They are interpenetrating; both are random; they give two independent estimates of the same thing. From each you can calculate what Dr. Deming calls the internal error. Then you have the difference between the estimates based on two independent samples which you can compare with the 'internal error' of the difference. This gives a most excellent check whether or not the sampling was done under statistical control.

Dr. Jessen:

I will take this opportunity to supplement a few of the ideas that were presented in the papers, rather than to offer any critical evaluation, since I am essentially in agreement with most of those ideas.

At Ames, we are conducting some research in sampling theory quite closely allied to sampling practice, inasmuch as most of our research is directed at problems met in making sampling surveys. We are at present engaged with the problem of the economics of sampling. We think a considerable amount of work needs to be done on this problem. As mentioned by Mr. Hansen, a sample design cannot really be made optimum until we know what the costs of the sampling operation are. Some of the results that appear to be coming out of our work indicate that a considerable amount of geographic distribution is usually advisable in most sample surveys, even though this means a considerable amount of outlay for travel and detailed mapping and other facilities required to specify such sampling.

On the matter of stratification, we have a little different philosophy towards that than that expressed by Mr. Hansen, for example. It is essentially this: Those characteristics of the population that can be easily measured, we generally prefer to use for purposes of sampling efficiency, in the probability of choice of sampling unit, or in the estimation procedure, rather than in the stratification operation; and to reserve for stratification those characteristics that are not so easily measured, like geography or soil types, or climate. By this means we think we can make more effective use of all the prior information that is available to us.

On the matter of area sampling, I notice that the idea is sometimes put forward that it requires a considerable amount of capital equipment. It is true that a considerable amount of money was spent in this country in the preparation of the master sample materials that are being used for a number of the samplings, by both the government agencies and nongovernment groups. This may mislead other countries into thinking they cannot carry out effective area sampling schemes because they don't have this equipment. We have got to the point now where we can do some of the mapping on a sampling basis that we did for every county in the United States and can do this much more effectively than we thought was possible several years ago. An example of this is a recent sample that we conducted in Greece, where certainly we did not have all the elaborate capital equipment for making area samplings. By this, I mean the mapping for sampling can be done for those particular places and for those particular blocks that come into the sample. This can be done rather quickly, and effectively, provided the proper kind of training is given to the interviewers, together with the purpose for doing these operations. They are inclined to think that this work is entirely irrelevant to the study, that their job is primarily one of conducting an interview, and not that of making maps and counts.

Dr. Chandra Sekar:

The All India Institute of Hygiene and Public Health, Calcutta, has recently been investigating the possibility of applying the technique of sample surveys for the estimation of birth and death rates. The following scheme was tried in their experimental center covering 68 villages within an area of 33 square miles and with a population of 63,000. An investigator had to call on each household, enumerate the population according to community, age, sex, marital and residential status, and list the births and deaths which occurred inside the village in the family during each of the last two years. An inspector who was in charge of a team of four investigators had to compare the births and deaths in the village as entered by the investigators with those found in the birth and death registration books and after field verification wherever necessary, work out the data into the following heads.

1. Number of events recorded correctly by the investigator and the registrar $= C$,
2. Number of events recorded correctly by the investigator only $= N_1$,
3. Number of events recorded correctly by the registrar only $= N_2$,

A supervisor was in charge of the four inspectors and sixteen investigators employed for the enquiry. Except for the supervisor, local residents were used as field staff as they are likely to obtain a better response from the people. Transport and housing difficulties were also thus obviated.

The field work lasted for eleven weeks and the per capita cost of the inquiry was approximately 1-1/3 *annas* or 2.7 cents per capita of the surveyed population.

For assessing the birth and death rates, the survey provides directly the population figure. The total number of births or deaths N is obtained by the formula

$$C + N_1 + N_2 + \left(\frac{N_1 N_2}{C}\right)$$

which takes into account the events omitted by both the investigator and the registrar. This estimate differs only slightly from "the maximum likelihood" estimate but is easier to calculate. The formula assumes that there is no association between the events missed by both persons. Special situations may be conceived when this assumption might *not hold good but such are not likely to be frequent under Indian conditions. Investigations are also in progress to assess the reasonableness of the assumption and to find out methods to remedy the situation when it fails.*

The variance of the estimate N is approximately equal to $\left(\frac{Nq_1q_2}{p_1p_2}\right)$ where $p_1 + q_1 = p_2 + q_2 = 1$ and p_1 and p_2 are respectively the chance for the investigator or the registrar detecting an event.

In the above inquiry the investigator's efficiency was not of a high order. 67 percent of the births of the previous year and 53 percent of births of the year before that were recorded by the investigators. 51 percent of the deaths in the previous year and 32 percent of the deaths in the year before that were detected by them. With greater lapse of time, there was greater difficulty in tracing the event. Deaths amongst persons aged below 10 years of age were missed more frequently than deaths of persons aged more than 10 years. Male deaths were detected more frequently than female deaths. The above results emphasize that field work will not usually yield accurate data and that suitable techniques have to be incorporated into the inquiry if it is to yield unbiased results.

For various reasons the entire population of the Institute's experimental center was covered in the inquiry described above. Another investigation undertaken by the Institute into the nature of correlation between the birth or death rates of adjoining geographical areas showed that to estimate the birth or death rate of a district with 1½ to 2 million population with a standard error of 5 percent, it will be sufficient to survey only about 6 percent of the population, selected by choosing at random the required number of sample units of 10,000 individuals living in contiguous villages. It thus appears feasible to estimate both the birth and death rates with fair accuracy at the low cost of 1 *pie* per capita or 0.2 cents per capita.

To conclude, the technique of sampling appears to be very promising for the estimation of birth and death rates in countries where theroutine machinery has still to be improved to yield data of sufficient accuracy. But the nature of the difficulties that impair the accuracy of the data collected in the field and the methods to overcome or allow for them have to be more clearly understood before this technique could be efficiently used in practice.

Dr. Hoijer:

In Sweden we have used for the population census of 1945 a method of sampling which may perhaps be of interest. The fundamental data about sex, age, family circumstances, profession, etc., we gathered from the whole population. We wished to have some additional information about education, employment, and national income; and in order to save time and money, we collected this information only from a part of the population. I should like to say some words about the method of sampling for this purpose.

Fundamentally, we chose all persons who were born in the month of July. From them the additional information was obtained at the same time that the complete-enumeration was made. By advertising and broadcasting, the public was informed that the July-born part of the population would receive these special schedules. But it was not possible always to isolate the persons born in July. For the existing families would have been divided; and it was essential to have the members of the family together. For this reason, the principle of the sampling had to be in some degree changed. If in a family, the oldest member was born in July, the whole family was selected to answer the special schedule. If, on the other hand, in a family the oldest member was not born in July, but there was another member born in that month, this family was not in the sample. This way, we had drawn a sample amounting to one-twelfth of the total population, or about 550,000 persons, spread over the whole country and giving a true image, I think, of the whole population, in respect to this distribution on sex, age, income, and so on.

Of course, we could have done the sampling in another way, too, by picking from the population lists of the total census every tenth or twelfth person. But then the sampling would have to be made after the total census; and we should have had to ask the sampled person twice.

Mr. Sukhatme:

I wish to mention a type of sample survey, extensively carried out in India, in which it is our aim to utilize the existing field agencies, each working within its own jurisdiction, and in which, therefore, the principle of maximum information per dollar plays a relatively minor role.

I have in mind the problem of estimating the average yield per acre. My country is divided into a large number of states, nine or ten; each state into districts, a district being approximately 4,000 square miles; a district into sub-divisions; and sub-divisions into what are known as *tehsils*. The ultimate unit of administration is a village (a little more than one square mile in area). In each village, there is an official whose principal duty is to collect land revenues; but who also maintains a register showing the names of contributors, the area in each field, the name of the crop grown, and so on.

This framework of administration determines the scheme of sampling almost automatically.

In the scheme of sampling, a village is a primary unit of sampling and a field within the village is the sub-unit. If we had a unit of sampling larger than the village, there would probably be a gain in precision. But we cannot adopt it because a village is the practical unit. With a larger unit, listings of fields would not be readily available.

The broad structure of sampling: namely, the primary unit of sampling, the field that comes into the sampling; and within a field, a small area from which the crop is to be cut are practically determined from these considerations of administration. What is determined from experimentation is only the number of sampling and sub-sampling units and their distribution.

But here again, travel costs do not enter into account, because it is one of the normal duties of the existing staff to carry out these experiments. What we do is to assist in applying the principles of random sampling under the existing conditions.

By adopting a scheme like this, we overlook certain theoretical principles of giving equal chance to every bit of land in the population. But these deviations do not seem to affect the estimate of final yield; and if they do, we can correct for these deviations by weighting the average yield in proportion to the area in each village, or similarly by weighting the average yield from each field by the area of that field.

One important point is the possible error arising from the human factor. I have especially in mind the size of the unit of sampling that we can harvest from a field.

It is well known that yields from neighboring areas are closely correlated. But with a small sampling unit, such as one would like to take, it has been observed that we overestimate the yield, sometimes by as much as 20%. This point is important because we must provide the existing staff with methods by which we can depend upon them to carry out the work so as to give precise results.

Dr. Molina:

There are one or two advantages in being old; in connection with the history of sampling, I am in a position to give Dr. Stephan the names of two men who did outstanding work many years ago, work of which I had first-hand knowledge.

I will first mention the name of an outstanding British telephone engineer, Captain G. F. O'Dell of the British Post Office. O'Dell made very extensive sampling studies of the distribution of traffic, that is, of the way telephone calls fall with reference to time and space. The result of this work was material changes in British telephone practice. If Doctors Kendall and Yates or any other of our distinguished English guests would like to learn more of the work of O'Dell, it is no use appealing to

O'Dell. He would disclaim any knowledge whatever of the subject. But they can obtain information through Egon Pearson, who is somewhat acquainted with the work of Captain O'Dell.

The other name is that of a distinguished American, the late Malcolm Rorty, a former president of the American Statistical Association. Rorty wrote his pioneer paper on the application of probability theory to telephone problems in October 1903. Basing our work on his pioneer paper, in 1905 we applied the theory of probability to the design of automatic telephone systems. This led to our invention of "translation of dialed pulses," an invention which revolutionized the design of dial systems.

But perhaps a more important result of Rorty's pioneer work, the impetus to which was given by sampling studies, was that he created a friendly atmosphere in the Bell system toward probability, an atmosphere which received with open arms the distinguished gentleman who introduced into the Bell System modern statistical methods and the art of quality control, Dr. Walter Shewhart.

M. Vincent:

Puisque M. Stephan fait appel á la collaboration de tous pour compléter la documentation qu'il possède sur l'histoire des applications des méthodes modernes d'échantillonage, je me permettrai de lui donner quelques indications recueillies à l'occasion de recherches effectuées dans le domaine de ma spécialité, c'est-à-dire en démographie.[1] Mais auparavant, je voudrais dire combien nous devons être reconnaissants à M. Stephan d'avoir entrepris un travail aussi ingrat.

Sur les premières évaluations de la population britannique, je me bornerai à signaler l'article que M. Glass a consacré à "*Gregory King and the Population of England and Wales at the End of the Seventeenth Century.*"[2] La statistique sur laquelle était basée l'évaluation de Gregory King avait été établie à la suite d'un *Act* de 1694. A peu près exactement à la même époque, en 1693, un "dénombrement" de la population française avait été préscrit; mais c'est seulement la grande enquête ordonnée en 1697, et qui donna lieu à la rédaction des fameux *Memoires* des intendants pour l'instruction du duc de Bourgogne (1697-1700), qui fut exploitée pour de nombreuses estimations de la population française, la première ayant été faite par Vauban et publiée dans son *Project d'une dîme royale*, en 1707.

M. Stephan a fort justement remarqué que ces premières évaluations de la population constituaient de véritables sondages. Les relevés de

---

[1] *Cf.* Paul E. Vincent, "French Demography in the Eighteenth Century," *Population Studies*, Vol. 1, No. 1 (1947).

[2] *Eugenics Review*, January 1946, p. 170 *ff*.

baptêmes dans les registres paroissiaux n'étaient souvent établis que pour fournir une base à l'estimation de la population, et la détermination du rapport existant entre les naissances et la population, dans des aires géographiques restreintes, n'avait pas pour but de fournir un indice de la situation démographique—comme, de nos jours, le calcul du taux de natalité—mais une évaluation du facteur de proportionnalité à adopter pour l'estimation du chiffre de la population. A cet égard, le sens dans lequel on prenait le rapport (25 habitants pour une naissance), inverse de celui adopté de nos jours, est d'ailleurs symptomatique.

La technique s'affinant, on en vint à effectuer de véritables sondages stratifiés adoptant des valeurs différentes du facteur de proportionnalité pour les villes et les campagnes, et suivant les régions géographiques. Un exemple de l'application de ces méthodes se trouve dans le remarquable ouvrage de Montyon intitulé *Recherches et considerations sur la population de la France,* publié à Paris en 1778 sous le nom de Moheau, un des secrétaires de l'auteur. Mais, semblables à M. Jourdain, les démographes du dix-huitième siécle faisaient des sondages sans le savoir. Le premier qui paraisse en avoir eu conscience est Condorcet qui, se livrant à "quelques réflexions générales sur la méthode employée par M. Moheau" dans le *Mercure de France* du 5 juillet 1778, p. 35 et suiv., posait la question suivante: "Je suppose par exemple que l'on ait opéré sur un canton qui contient 200,000 personnes, qui l'on ait observé que le rapport du nombre des naissances à celui des individus est comme 1 à 23, et qu'on en conclue que dans un pays où il nait un million d'enfants par an la population sera de vingt-trois millions; je demande maintenant jusqu'à quel point je puis compter sur cette conclusion." Et Condorcet exprimait son espoir que le moment viendrait où les travaux des probabilistes deviendraient utiles, grace à "un homme né avec le génie des Mathématiques et celui de la Philosophie."

Celui qui devait fournir, cinq ans plus tard, la réponse à la question posée par Condorcet, fut son collegue à l'Académie des sciences, Laplace; et, si l'on songe aux relations étroites qui s'étaient établies entre eux, précisément en ce qui concerne les questions démographiques—ne devaient ils pas, à partir de 1783, présenter en commun à l'Académie, associés à Du Sejour, l'*Essai* de La Michodière sur la densité rurale—il n'est pas exclu que la question de Condorcet ait été à l'origine des travaux qui devaient conduire Laplace à la première application du calcul des probabilités aux problèmes soulevés par les sondages démographiques (*Histoire de l'Academie royale des sciences—annee* 1783-.., pp. 693-702).

Mr. Shafei:

I would like to remark on Mr. Hansen's method of sampling of blocks. I feel that in relying on the first numbering of the blocks, (I presume

that the numbering of series of blocks would proceed probably geographically) and in taking every n'th block this would not represent the population exactly unless the density of the population is even over the area. There are some blocks, as we know, that tower high up; and therefore, they contain so many inmates; and therefore, the n'th block in the cluster of an urban geographical area will contain many more inhabitants than, for instance, the $(n+4)$th block. And if the blocks or the streets are ordered alphabetically, there is a danger in some languages that the several letters will not have the same frequency of utterance. For instance, if you look at a telephone directory of an Arabic-speaking country like Egypt, you will find the letter M, for instance, is very much more frequent than any other, the reason being that the name Mohammed (which relates to Islam) is very much more frequent than other names. Therefore, if we depend on alphabetical ordering, we run into bias without being conscious of it.

I would like to suggest a method of choosing areas, random areas. Supposing we have an area or a city with a map of, say, ten inches by ten inches square; and suppose we want, say, a ten percent sample. I would like to suggest a random square of one inch side spread randomly over this area. In order to determine the position of this square, use horizontal and vertical coordinates. These coordinates will be determined randomly from the figures one to ten. If I have the horizontal axis ten inches long and the vertical axis ten inches long, I shall have ten coordinates, numbered from nought to nine on the vertical side, of which I can choose any number at random. And, therefore, I can assign any number of these sample squares. If I want a ten percent sample I can have ten such squares by choosing ten pairs of random coordinates. And this will give me the ten percent sample spread over the area of the town in a very random way. This will also, I believe, give me a chance of doing away with stratification of the sample. Wherever I take the one-inch block, I can make the interviews within that block. And if I make the interviews in the ten blocks then I shall have a ten percent interview, or a ten percent sample. I think this will be a random one.

Dr. Yates:

I have to thank you, Madam Chairman, for giving me the opportunity of making a few remarks. I was speaking yesterday on theory and I don't want to say much today, but I would like to mention one or two items of work with which I have been concerned on the practical side.

First I should like to mention the Survey of Fertilizer Practice. The reason I want to mention it is that it is a good example of what may be called very untidy sampling. This untidiness was forced on us by the nature of the material. Had we attempted to be orderly we should have failed to collect the information that we required.

The Survey of Fertilizer Practice was designed to find out how farmers were using their fertilizers on the different crops. The information was required in considerable technical detail. In addition, soil samples were to be taken from soil of the fields for analysis for fertilizer requirements. The sampling procedure we adopted was as follows: a sample of farms was first taken, using a variable sampling fraction so that more of the larger farms were secured. This part of the sampling is very simple in England since we were able to obtain addressograph lists of farms in each county which contained a record of the acreage of each farm. The field workers visited the selected farms and took a sample of one field of each crop growing on old arable land, a similar field of each crop growing on new arable land, (*i.e.*, land ploughed up since the war) and one or more fields of permanent grass. This part of the sampling was done in conversation with the farmer, making a list of all the fields classified by the different crops, etc.

As you will see this results in sampling fractions which are different for almost every unit in the sample. This results in a good deal of computing work in the abstraction of the information. This is an inevitable necessity because had we asked for information from the farmer for all the fields on his farm this would unquestionably have put too much strain on him, and had we taken a random sample of all the fields on the farm we should not have had a good coverage of each crop and would also have irritated the farmer by, for example, asking him for information on two or three of his wheat fields and nothing about his sugar beets.

I would also like to mention an example of how biological and sociological problems interact in sampling. During the war we had the problem of determining the number of people that were about in a given area of a town in order to see whether air raid shelter facilities were adequate. The problem was of no great importance in that it turned out the information was never seriously required, but given the problem the solution is interesting and is relevant in problems of ecological survey.

The police authorities, when confronted with this problem, always attempted to solve it by posting observers at fixed points with instructions to count passers-by. Now it is clear that such counts will not give an answer to the number of people in an area unless all exits and entries can be covered and the counts are continuous and accurate. In fact, these conditions cannot be even approximately fulfilled in the case I have mentioned. Instead of using stationary observers, therefore, we proposed that moving observers should be used. Thus, for example, in order to find the number of people in a street one or more observers (depending on the extent of the crowd) proceeded down the street counting all the people that they passed and deducting anyone who passed them. At first sight this sounds a difficult operation but in practice it proved to be quite easy. In streets with few people similar counts can be made from moving vehicles. The counts must be made in both directions in order to allow for any predominant flow in one direction. A similar procedure can be used to find out the number of people in a store. It was actually used for this purpose and proved far more convenient and accurate than previous attempts based on counts of people entering and leaving the store.

Dr. Szulc:

One of the earliest applications of scientifically developed sampling methods to the tabulation of the population census took place in Poland in 1932. Material from the population census of December 9, 1931, was available; but they could not wait until the complete tabulations would be ready. And they needed cross-tabulations not ordinarily included in census tabulations. The information needed was the population by number, age, sex, family condition, age of wife, number and age of children, and so forth.

The sampling method was designed by Professor Neyman. Most of the modern sampling methods, stratification, sub-sampling, double sampling, were used in this investigation. Some theoretical results of the design of this investigation were developed by Professor Neyman in a paper on random and purposive sampling published in 1934 in the *Journal of the Royal Statistical Society.*

Miss Walker:

Prof. Mahalanobis sends up a note calling attention to the fact that the method of squares located at random, with the help of random coordinates, which was mentioned by Dr. Shafei from Egypt, has been used in India as a routine measure since 1937.

Dr. Sastry:

In the examples of the uses of modern sampling procedures given by the three speakers, I miss any reference to the application to the affairs of banks. This may, perhaps, be due to the fact that bankers in general are very conservative, and do not like innovations which do not give accurate results, but only estimates. Recently modern sampling procedures were applied in certain inspections of banks in India; and interesting results were obtained.

From statutory obligations, the Reserve Bank of India has to estimate the available assets of certain member banks called stable banks. A member bank has usually a very large number of branches spread over different regions of the country. Each bank has transactions with a large number of constituents; in these loans, we have different collateral securities: for example, government bonds, equity shares, real property, and in some cases personal demand notes. The value of these securities at the time of inspection will be different from those at the time when the loan was given. No doubt, the bank will have the usual margins when lending the amount. But given a financial crisis, the margin will not be sufficient; hence, the object of the inspection is to estimate the total available amount of all these loans. The usual procedure is for the inspecting officer to select a number of doubtful branches; and in these offices to check loans of a particular amount.

The available amount among these loans is calculated and the identical amount is given to the uninspected bankers and unexamined loans. Obviously, this method can lead to serious bias in estimates of available assets.

The new sampling procedure is as follows: First of all, for each bank, from the total number of branches a few branches have to be selected; and in each branch, of all the given loans, a certain proportion of loans will be selected at random. I am not going into the details of all this procedure; but I can only say that the estimates found are within the limits of probable error.

Mr. Stephan:

I have been gratified by the suggestions that have been made and hope that others may send me additional examples that would be helpful in this history.

I wish that there had been enough time to cover the developments of the last ten years. There are some broad and important trends; Dr. Jessen has referred to the development of the economics of sampling. There were several early attempts to deal with problems of cost and to find the optimum design of a sampling operation from the standpoint of the cost system. A great deal of important work is now going on, and more will be done in the future, on that problem. It is not only a problem of capital investment and current cost, but one of sharing the cost of preparatory work among many surveys so that it may become possible to do sampling as part of a broad scheme of sampling projects, when such sampling would be too costly if undertaken by itself. That is one reason why the development of sampling centers at the Census Bureau, Bureau of Agricultural Economics, and Bureau of Labor Statistics, and many of the universities and private commercial research organizations is a great step forward. It means a great deal for research to have accurate and economical methods of sampling available.

There is another trend that we see only dimly now, and that is increasing recognition of the problem of explaining to the public how dependable a sample is, and when it is safe to treat sample data as if they were absolutely accurate. We have not yet solved the problem of training millions of people to know what a standard error is, or fiducial limits, or even to recognize that a set of sample data may be dangerous for some purposes and safe and valuable for others. In this problem of informing the public, statistical organizations have a great function to perform.

Some reference was made to the training of interviewers and of other people who collect the sample, in accurate methods of sampling. This is another important problem on which work is to be done; and it calls for psychological research of a high order. Coupled with it is the psychological problem of eliciting from respondents accurate answers to questions. We have before us vast and difficult problems of improving the accuracy of our methods of measurement. As those problems are solved, the advantages of accurate sampling methods will become more important, and more valuable.

We are also making additional use of accumulations of data, both in the design of samples and in the preparation of estimates from them. That calls for attention, too, and shows again a link between practice and sampling theory.

Finally we are bringing together the various aspects of sampling theory and sampling practice into a systematized body of knowledge. In other words, we are dealing with a complete system rather than with a lot of separate sampling techniques.

As to the specific comments, I was glad to have Dr. Molina's references to the early work in the American and British telephone systems. The fact that this early work was not given greater publicity was partly due to modesty, and partly due to the fact that the men who made these excursions into sampling were preoccupied with other problems. At the same time there was an interesting discussion of sampling methods, including considerations of cost and the development of an optimum sample design, among the mining engineers. They were stimulated by the problems in the western mines and South Africa, of estimating the gold content of ore, and of controlling the mining processes so that any leak by which the gold was being diverted might be discovered. There were also problems of either deliberate or unconscious contamination of samples; called "salting." We sometimes have those problems today in economic and social sampling.

I am glad to have the comments that Dr. Vincent gave us, from the early history of demography. I was not unaware of some of these developments, but did not have space to refer to all of them. Laplace's contribution was unique in one way: he followed methods that others had followed in preparing population estimates but, so far as I know, he was the first to put limits on his estimates and to give the odds that he might be wrong by more than a certain amount. It is to be regretted that this lead was not followed up, because these intellectuals and mathematicians, predecessors of modern statistics and demography, could have developed a great part of modern sampling practice and the theory that underlies it, if they had been compelled by practical problems to turn their attention to it. Here again, we find the interaction of practice and theory, exhibited this time by lack of sufficient practical pressure on the men who were capable of developing the science.

I am grateful to Dr. Szulc for his references to Dr. Neyman's papers, and to the work that was done in Poland. Dr. Neyman has stimulated development in sampling theory and practice in a number of important directions.

Several similar tabulations of complete census data have been made to obtain cross-classifications that would not otherwise be available, notably in England and the United States. In the census of 1940, we used a five percent sample taken simultaneously with the enumeration, to get additional information about the population. But we also used that sample to get detailed tabulations quickly, and with adequate cross-tabulations, for information that was available for the whole population. It was economical, but even more important, some figures were available six, eight, or twelve months in advance of the figures that were provided by the complete tabulations, simply because sampling methods were used. When the complete tabulations were made, the sampling system proved itself by showing very small errors. The few outstanding errors that did appear were traceable to changes in the coding procedures, between the coding of the sample material and the coding of the complete census. They called attention to the fact that other sources of error are frequently more important than the sampling error of a well-designed sampling procedure.

Mr. Hansen:

Mr. Sukhatme mentioned that, where he had an existent field organization available to work with, the best sampling system did not involve necessarily the application of the theory of optimum sample design. I should like to emphasize what I referred to in my original statement: that we do not have any theory of optimum design, such that there is a set of formulas in which we can put certain numbers and uniquely determine the optimum. What you can do, and I think Mr. Sukhatme's problem comes within that framework, is make a decision as to a type of sampling approach. This decision should take account of the resources and facilities that are available. Then within the framework of this broad approach sampling theory will guide one to the optimum. The field organization available is one of the resources to be utilized, and in

Sukhatme's case the theory should point to an optimum design that makes effective use of this field resource.

Now, with respect to the remark by Mr. Shafei, he indicated that a bias would result from using the methods that I had suggested if the population was not distributed uniformly within a city; and that if one sampled random squares he might avoid this bias. The fact actually is that neither method would be biased. Either method, properly used, will give unbiased results, so that the decision between one type of approach and the other would be based on the information available for estimation, stratification, and other purposes, and which way you could carry out the various field enumeration procedures. Very often in block sampling the use of street boundaries has very substantial advantages.

One other thing I wasn't quite sure of in Mr. Shafei's remarks was the suggestion that you might take a ten percent sample of a city by taking ten random areas from a group of 100. That would be a random sample, true enough, but if you took only ten and used such large areas that you had only 100 such areas in your grid and then took ten at random, the sample would ordinarily—even though a high percentage of the population, perhaps ten percent, was included in the sample—be based on such large clusters that it would be very inefficient and estimates from such a sample would have a large variance. But the method when used with small-sized clusters would be a competitor with other methods, and the one to be used would depend upon the circumstances and available facilities.

Dr. Hauser:

Madam Chairman, if I may be permitted two very brief observations, from the standpoint of one who has been strategically located for participating in, and observing a large number of social surveys. It has occurred to me particularly, as I have listened to the sampling papers, that one must leave with the notion that probably no survey ought ever to be approached without first raising the question of how large a sample is adequate for the purposes of the given survey. Resources are frequently wasted, through distrust of, or lack of familiarity with the efficient sampling methods which are available. I am sure that the resources now going into the collection and dissemination of statistics would make available a much broader range and scope of data to the government administrator, legislative bodies, business, labor, and other user of data if efficient sampling methods were utilized. The point to this observation is that I think that each of us engaged in the collection and dissemination of statistics, were we to ask the question with respect to every survey, of how large a sample is adequate to the purpose, and use the efficient forms of sampling design which are available, we could provide our consumers with considerably more data for a given expenditure than is now actually the case. This is a lesson flowing from these meetings that I think is worth pondering.

I have a second observation, if I may be permitted to make it. It is that in a social survey, that aspect of methodology which has become the easiest to solve is that of sampling. We should be grateful for what has been developed in this area as an important contribution to adequate and effective results. But we must also be mindful of the great areas of ignorance, which remain in many other aspects of the social survey, and the importance of exercising the same ingenuity as is evident in the development of sampling techniques, to those other sources of now unknown and, in the main, unmeasurable error in the collection of social and economic statistics.

DEVELOPMENTS IN SUBSTANTIVE ECONOMIC AND SOCIAL STATISTICS IN THE UNITED STATES

*Saturday, September 13, at 2:00 p.m.**

CHAIRMAN:
Isador Lubin

RECENT DEVELOPMENT IN MARKETING STATISTICS

by Ross M. Cunningham
Massachusetts Institute of Technology
(United States)

Marketing includes the various business activities involved in the transfer of goods and services from producer to consumer. The present analysis will be limited to a consideration of marketing from the point of view of the manufacturer of consumer goods, or products which are used by the general public.

Marketing research is the scientific study of marketing problems and makes widespread application of statistical methods. It has pioneered quantitative measurement in an increasing number of marketing activities where opinion and hunch once prevailed, but by and large has not been concerned with development of statistical theory. Hence, the discussion of recent developments in marketing statistics will treat of (1) the general growth of marketing research and its organization within industry, and (2) significant advances in marketing research techniques and opportunities for further advances.

I. GROWTH AND ORGANIZATION OF MARKETING RESEARCH IN INDUSTRY

The substantial growth in use of marketing research by industry in the past fifteen years is common knowledge to marketing men but there is no fully satisfactory quantitative measure of this growth. The first comprehensive study of manufacturers' marketing research activities was made for the calendar year 1937 by the Department of Commerce in cooperation with the American Marketing Association.[1] This inquiry yielded much new and valuable information but was directed to a relatively small selected group of companies and did not purport to be a census. In 1945 and 1946 two further studies of marketing research within industry were made, one by the National Industrial Conference Board[2] and one by the American Marketing Association in cooperation with the National Association of Manufacturers.[3] The latter study, although limited to members of the National Association of Manufacturers, represented a large enough response to provide an approximate indication of the present use of marketing research among manufacturers. Of the 4,786 companies replying out of 13,000 on the mailing list, 38% carried on some marketing research either through an organized department or through a line executive. An organized department was defined as one or more people engaged in marketing research. Of the 1,485 manufacturers of consumer goods, 15.1% had marketing research departments and an additional 31.3% conducted research through a line executive.

---

* Meeting arranged by the American Statistical Association.

These three studies, while providing invaluable data on the conduct of marketing research, do not furnish a satisfactory basis for measuring growth of the activity. There are, however, other indications of growth, such as the larger volume of literature, the increase in number and size of specialized marketing research companies, greater interest and activity by various professional and trade organizations and expansion of the membership in the American Marketing Association. The last is a particularly useful measure as one of the basic objectives of this Association is "to advance the science of marketing" and members either are interested in the analytical approach to marketing problems or are working directly in the marketing research field. Membership now stands at nearly 3,000, a five-fold increase since its formation in 1937 as a merger of two predecessor associations, the American Marketing Society and the National Association of Marketing Teachers.

---

[1] Elma S. Moulton, "Marketing Research Activities of Manufacturers," Market Research Series No. 21, U. S. Department of Commerce, 1939. The American Marketing Association suggested that the study be undertaken and appointed an advisory committee to help in formulating plans.

[2] G. Clark Thompson, "Organization for Market Research: Part I, Industry Experience; Part II, Operating Methods and Company Plans," *Studies in Business Policy*, Nos. 12 and 19, National Industrial Conference Board, 1945 and 1946.

[3] Heusner, Dooley, Hughes. and White, "Marketing Research in American Industry," *Journal of Marketing*, Vol. 11 (1947), p. 338; and Vol. 12 (1947), p. 25.

Thus, there has been important general growth in use of marketing research by manufacturers of consumer goods and generally by all manufacturers. There is some evidence that industrial goods producers are accepting marketing research at an accelerated rate.

Space does not permit detailed treatment of the organization of marketing research in industry. Some few conclusions based on the aforementioned studies may be of interest, however, and are stated without discussion.

1. Consumer goods manufacturers are more likely to conduct marketing research than industrial goods manufacturers.
2. A larger proportion of consumer goods manufacturers have an organized marketing research department than do manufacturers of industrial goods.
3. Large companies make greater use of marketing research than small companies. There are relatively few separate departments in manufacturers with annual sales under \$2,500,000.
4. Top management is actively interested in marketing research. Where there are separate departments about three out of four report to a top level executive.
5. Many types of marketing research are carried on but major emphasis seems to be placed on studies of consumer markets, of product lines, of distribution methods, and of advertising and other sales promotion methods. Substantially similar interests were found among manufacturers of industrial goods.
6. The cost of operating marketing research departments for the companies reporting such information was 0.3% of net sales in the two recent studies. This represents a very substantial increase over the rough measure provided by the 1937 study.
7. Manufacturers with marketing research departments stated that they intended to expand the activities of these departments.

II. RECENT DEVELOPMENTS IN MARKETING RESEARCH

It is next in order to consider briefly the specific applications of marketing research which have shown particular development over the past 15 years and to point out opportunities for further development. The first step here is to list the major marketing fields in which a manufacturer of consumer goods might conduct research studies.

(a) The Consumer Market
(b) Studies of Manufacturer's Product Line
(c) Trade Channels and Distribution Policy
(d) Price Levels and Price Policies
(e) Sales Promotion Methods

(f) Physical Distribution of Products
(g) Financing of Marketing Activities
(h) Public Relations and Governmental Relations
(i) Planning of Marketing Activities

A review of marketing research developments in terms of this outline reveals that they tend to be heavily concentrated in two general fields—the Consumer Market and Sales Promotion Methods. These accordingly will be considered first and the others treated very briefly afterwards.

The Consumer Market

Studies of the consumer market are of substantial importance to a manufacturer because he has no direct contact with buyers and users of his product in the normal course of his business. Consumer goods typically flow through trade channels, both wholesale and retail, which effectively conceal consumers' buying behavior and attitudes. Furthermore, the measurement of flow of manufacturers' sales from the shipping platform is misleading as an index of market acceptance, for a very considerable time lag may exist between the shipment of goods to a wholesaler and the movement of those same goods into consumers' hands.

(1) *Increased Attention to Design of Samples:* Sampling is the most basic and widely used statistical tool of marketing research. Almost never is it practical for a manufacturer of consumer goods to make a census of his market.

The last few years have seen active and sometimes acrimonious discussion over the relative merits of "quota" sampling versus "area" sampling. Quota sampling as commonly practiced is a stratified proportional sample (though it may be non-proportional) in which interviewers are assigned quotas of the various strata desired. The interviewers may be instructed to work in certain territories but are usually allowed discretion in selecting interviewees so long as these conform to quota requirements.

The area sample is derived by dividing the market being studied into small geographical areas, selecting a sample of these at random, and within each such area selecting a random sample of individuals or households to be interviewed by appropriate sub-sample design. Call-backs are made until coverage is as complete as possible. Area sampling in its purest form does not require advance knowledge either of the population being investigated or of strata which may be relevant. Closely related to area sampling is the situation in which the population is known, and the units can be identified in a manner which permits selection of a random sample. Interviewers then are assigned lists of individuals or addresses. The inevitable geographical scatter of this sample design results in high travel costs.

It is not appropriate in this paper to review the arguments pro and con of quota and area sampling. Some of the discussion has certainly arisen through lack of sufficient understanding of the other side's point of view. Statisticians have deplored quota sampling yet have not always been close enough to day-to-day operation of surveys and to the usefulness of their results in actual practice to appreciate fully the importance of practical "know-how." On the other hand, practitioners have been too much influenced perhaps by higher costs of other sample designs, by the apparent workability of methods currently used, and by the difficulties of undertaking new methods.

This discussion appears to be subsiding as there is wider recognition that the sample design should be governed by the particular circumstances of the investigation and that the use of any single design for a wide variety of studies is likely to be either inaccurate or unnecessarily costly. The degree of knowledge of existing market behavior and the factors which cause this behavior is important in designing a sample. Stratification of a sample without proof of relevance of the strata selected will not reduce sampling error, except through sheer coincidence. The proportion of total households using a type of product and the variability in conditions of use between households are also important factors. A widely used product with essentially the same applications in each household can be studied successfully with a fairly simple sample design.

Furthermore, experimentation has revealed that modifications of area sampling methods can sometimes be introduced without serious loss of precision and with appreciable reductions in cost. How far this can be carried, of course, depends upon particular circumstances. One of the very real needs today in sampling is intensified analysis of the various sets of conditions under which different types of sample design are appropriate.

A second need is further study of applicability to marketing problems of the theory of sampling from non-symmetrical populations. Most of the statistical tests now used in measuring market samples are premised on a symmetrical distribution of the population, yet we know that many market populations are highly skewed in their distribution.

(2) *Measurement of Consumer Purchases by Brands:*

Techniques have been developed for measuring at regular intervals consumer purchases of product classes in total and by individual brands. The A. C. Nielsen Company in the early 1930's introduced a method of measuring sales by brands across the counters of a sample of retail stores located throughout the United States. The results were then projected to provide national totals. This service appeared first in the food field and later in the drug field. Stores are visited every sixty days by field representatives at which time physical inventories and invoices are checked to establish sales over the period. There are also some other somewhat similar commercial operations in particular areas and a few manufacturers have set up their own store audit arrangements.

Prior to the appearance of this type of analysis there had been spot studies of consumers and some use of so-called consumer juries for certain aspects of product testing, but there had been no means of providing information at regular time periods by which the sales of individual brands and of the total product class could be measured. This technique, of course, covers only product sales through the particular type of outlet included in the sample.

A second and somewhat different type of consumer purchase measurement is the panel of consumers who report their purchases of specified product classifications, usually on a monthly basis. This technique began to be important in the late thirties and early forties. Experimental panels have been operated by two of the leading advertising agencies, and some panels are operated commercially by marketing research organizations, such as Industrial Surveys Company, Inc., who sell the service to their clients. Manufacturers have also experimented with consumer panels.

Usually the data are secured by forms which are filled out by the panel families, but a certain amount of regular field contact is maintained in verifying socio-economic classes of existing panel members and in the selection of new families to be added to the panels.

Consumer panels cover purchases from all types of outlets, measure the extent of repeat purchases, measure rates of consumption, provide information on exposure to various advertising media, and furnish a substantial amount of data on the family unit. Continued development of the consumer panel technique is likely.

(3) *The "Why" of Consumer Buying Behavior:* Although new techniques provide much information on what consumers are buying, they do not reveal much as to the "why" of this buying behavior. In the last few years greater use has been made of psychologists and of psychological investigation techniques. In this general effort the depth or open-ended interview has proved useful, especially as an aid in formulating hypotheses for investigation. There is also at least one experimental consumer panel in which information of a psychological character has been gathered on each household. Expansion of this type of work will undoubtedly take place.

(4) *Questionnaire Phrasing:* Public opinion specialists have devoted considerable study to the effect of question wording on answers to public

opinion polls. Unfortunately, there has not been a corresponding amount of attention to question phrasing in the typical commercial investigation. This is a field in which further work would be desirable.

(5) *Selection, Training and Supervision of Field Interviewers:* The importance of competent field workers is self-evident, and the last few years have seen a considerable growth of interest in more effective selection, training, and supervision. The fact that these interviewers are generally part-time workers complicates the problem but progress is being made and efforts seem likely to be intensified.

Sales Promotion

Marketing research in this general field has been very active both because manufacturers are increasingly seeking maximum return on their sales promotion dollar and because advertising media have used research as an effective means of promoting their sales of space or time. Only a few of the more important types of studies can be mentioned. The Advertising Research Foundation, a nonprofit organization sponsored jointly by the Association of National Advertisers and the American Association of Advertising Agencies, has been instrumental in undertaking or advising in many important studies in the field of advertising.

Circulation measurements and studies of the economic and social composition of circulation have been frequent. One interesting analysis of circulation composition was carried out by the Curtis Publishing Company which gave the Census Bureau a sample of 10,000 households buying the Saturday Evening Post to match against the population cards of the 1940 census, which were then tabulated in various ways.

Under the general head of circulation measurements comes the Continuing Study of Magazine Audiences sponsored by Life Magazine.

The problem of measuring listenership to particular radio stations (as opposed to individual programs) is a difficult type of circulation measurement. The first comprehensive efforts to solve this was made by the Broadcast Measurement Bureau.

Outdoor advertising through the Traffic Audit Bureau developed during the 1930's methods of measuring exposure of people to outdoor posters. More recently a pilot plant study using new techniques was conducted in Fort Wayne, Indiana, to secure more precise data on circulation values of outdoor posters and to gain information on the effectiveness of this advertising.

Measurement of consumer exposure to individual advertisements or to editorial content has been developed in several media. In the case of magazines, the Daniel Starch organization conducts field interviews which review the magazine page by page, noting whether the advertisement has been seen, whether some of it, or whether all of it, has been read.

In measuring radio advertising, the Nielsen Audimeter, an electronic recording device attached to radios in a sample of homes keeps a continuous record of the times at which the set was turned on and the stations to which it was tuned. Because the device operates continuously, listenership can be determined throughout each minute of the program. The tape on the machine is changed monthly by field representatives who at the same visit make an inventory of the radio-advertised commodities found in the home.

The second major method of radio program measurement conducted by C. E. Hooper, Inc., is by telephone calls made at the time a program is on the air. A short series of questions determine whether the radio is turned on and the station or program to which it is tuned. A number of calls taken at random from a telephone book is made in each quarter hour and provide the sample for that time period. The sample is necessarily restricted to telephone homes. The entire field of radio measurement technique will see further changes and advances.

The American Newspaper Publishers Association through the Advertising Research Foundation has been conducting the Continuing Study of Newspaper Reading. Studies have now been completed in over 100 cities. These studies are spot studies and do not yield trend data for any particular city. They do, however, provide a measure of readership both of editorial matter and advertising.

Another study conducted by the Advertising Research Foundation is the Continuing Study of Transportation Advertising and is supported by the National Association of Transportation Advertising. These studies are being made in a series of representative cities and appear to be yielding much worthwhile data in a field which previously had limited quantitative measurements.

Point of sale advertising is beginning to be investigated more carefully, often through the cooperation of retail store organizations. Conditions of use of such advertising often will permit carefully controlled experimentation which approaches laboratory work more closely than can most of the marketing research measurement techniques. Opportunities appear good for fruitful results in this field.

There have been fewer studies in the sales management field than in advertising. However, there has been real progress in developing tests for selection of salesmen. Also there have been slow but steady advances in training methods, in establishing sales territories, and in setting sales quotas. In particular, there has been significant work done in time and duty analyses for wholesale drug salesmen.

Other Fields

In the remaining fields there have been few major developments and in some fields research activity is conspicuously absent.

Product studies have undoubtedly become more comprehensive and precise. Pre-testing is carried out carefully with typical consumer groups, and sales tests under normal marketing conditions also yield useful information.

In distribution policy many manufacturers have been considering greater selectivity among outlets and probably there may be many unpublished studies. A chief disappointment to students of trade channels is the failure of Congress to pass necessary legislation for a census of distribution. Business must come fully awake to its responsibilities in asking for government information which it wants and needs. In the absence of up-to-date information counts by some newspapers of numbers of retail outlets of various types in the shopping centers throughout a city will prove helpful.

Little research is going on in the field of pricing and this is one of the really important gaps. In particular, we need better ways of establishing prices for new products.

Little activity of significance is evident in physical transportation and financing fields, but it must be recognized that research work in these fields tends to be done by other agencies than marketing research departments of manufacturers. In public relations, however, marketing research has been of growing assistance. Also, the marketing research department has tended to play an increasing role in the planning of marketing activities.

Government and Other Published Data

Little mention has purposely been made of published data from governmental and other sources. It is clear that marketing research relies heavily upon such information and also clear that the usefulness, quantity, and currency of such data are increasing. The government agencies are aware of the needs of marketing men and have sought their counsel in planning for future activities. These programs will be covered by the following speakers and need not be treated here.

Résumé

L'analyse des marchés dans l'industrie manufacturière a pris une forte extension au cours des quinze dernières années et continuera à se développer. Les entrepreneurs la considèrent de plus en plus comme étant de très grande importance.

Les techniques de l'analyse des marchés ont montré leur plus grand développement dans des études concernant les marchés de consommation et les méthodes de développement de la vente.

En matière de marchés de consommation, les développements ont concerné principalement la procédure d'échantillonnage, la mesure des achats par marque et l'analyse de ces données selon d'autres facteurs. On procède à des enquêtes sur les aspects psychologiques des réactions des acheteurs. Il serait désirable d'étudier davantage la rédaction des questionnaires et la formation d'enquêteurs.

En matière de développement de la vente, la publicité a été le plus important domaine de recherches; des études ont aussi été consacrées à la circulation et sa composition et à la mesure dans laquelle les différente formes de publicité agissent sur le public. Des analyses concernant le temps et les quotas des vendeurs ont formé un développement notable dans le domaine de l'organisation de la vente ainsi qu'un plus grand usage de tests pour le choix de vendeurs, mais il est nécessaire d'étendre les recherches dans ce domaine.

Dans d'autres domaines: canaux de distribution, fixation des prix, distribution physique, financement, relations avec le public et le gouvernement, et la planification des marchés, l'analyse ne montre pas de développement d'ordre comparable. Les importants problèmes que présentent ces divers domaines prouvent qu'il est désirable d'intensifier les recherches. Il est particulièrement urgent de développer l'analyse en ce qui concerne la fixation des prix, spécialement de produits nouveaux. .

PROGRAM OF THE CENSUS BUREAU

by Philip M. Hauser
Deputy Director, Bureau of Census
(United States)

The statistical program of the Bureau of the Census has been under careful scrutiny by its administrative and professional staff during the past several years. The program, procedures, and management of the Bureau have been evaluated with a view to: (1) maximizing the usefulness of the Census product to consumers of the data; (2) improving and modernizing methods and techniques; (3) increasing administrative efficiency. Although many important substantive, methodological, and organizational changes have occurred in the Bureau of the Census in recent years, many of them stimulated or necessitated by the requirements of war, the task of designing a useful, reliable, balanced, and efficiently engineered product is a continuous process and is still actively under way.

The statistical work of the Bureau of the Census falls into six distinct fields which are the basis for the organization of six of the Bureau's twelve divisions. These fields are population (including the labor force and housing), agriculture, industry, distribution (including wholesale, retail, and service trades), government statistics, and foreign trade.

In all these fields, with the exception of foreign trade, the program can be considered in two distinct phases: (1) the collection and publication of complete census or benchmark information; (2) the collection and publication of current information. In the field of foreign trade, monthly statistics are collected and disseminated on a continuous basis.

In response to increasing needs for current information of Government Agencies at Federal, State, and local levels, of business, of labor, and virtually every other aspect of American life, increasing emphasis has been placed upon current statistics. In fact, with respect to most of the program of the Census Bureau, it may be said that emphasis has been shifted from complete census statistics largely for the historical record, to current statistics for current use. This shift in emphasis has of course placed a great premiun on speed and clarity in the collection and presentation of data. It has become a major objective of the Bureau to make statistics available as soon as possible after the date to which they apply, and in a form in which the contents may be easily comprehended.

THE SUBSTANTIVE PROGRAM

The substantive program of the Census Bureau can be considered in two parts: first, general demography including population and agricultural data; and second, general business and economic statistics. These categories are not mutually exclusive but they provide an effective basis for administering the work of the Bureau.

At the present time, under the law, a complete population census is taken decennially in the years ending in "0" and a complete census of agriculture quinquennially in the years ending in "0" and "5".

In recent years, important developments have occurred within the Bureau in the preparation of current population information including information on the labor force and housing in addition to the inter-censal and post-censal population estimates which have been prepared over a long period of time. A sample survey of the population of the United States is being taken on a monthly basis to provide, for the nation as a whole, current information on persons in the labor force including the employed and the unemployed and their general characteristics, such as age, sex, occupation, industrial affiliation, class of worker, hours of work, and duration of unemployment. This same sample survey also furnishes current information on the general characteristics of the population including family composition, school enrollment, and migration, and on the characteristics of housing. The compilation of such information on a monthly basis through a canvass of the population has become feasible only as a result of recent developments in the theory and practice of sampling human populations. The types of statistics indicated are based on a small sample of 25,000 households selected to represent the entire population of the United States.

In the field of agriculture, current statistics on crops, livestocks, and characteristics of farms and farm population are obtained primarily through the Department of Agriculture. The Census Bureau, however, does collect current data on cotton ginning and fats and oils.

Plans for improving the usefulness of statistics in the field of population and agriculture include the possibility of a quinquennial census of population to be taken simultaneously with the quinquennial census of agriculture. (A simultaneous census of population and agriculture occurs now only in the years ending in "0"). Also being considered are an annual sample census of population and agriculture in the intervening period which would be adequate to report population, labor force, and housing data for states, important cities, and metropolitan areas, urban, rural farm, and rural nonfarm population of the states, and the population of smaller cities by city size groups. The proposed annual sample census of agriculture would provide data on crops, livestock, characteristics of farms, and farm population by regions and states.

In the general business and economic field, there is conducted at the present time a decennial census of distribution (including the wholesale, retail, and service trades) for the years ending in "9"; census of mineral industries for the years ending in "9"; and a census of state and local governments for the years ending in "2". The census of manufactures is authorized biennially, and is taken to cover industrial activity in the odd years. Experience has shown that this program is not adequately balanced in the light of consumer need, and there is now pending in the Congress of the United States new legislation which would authorize a quinquennial census of distribution, manufacturing and mineral industries and also a new census of transportation, excluding railroads (for the years ending in "7" and "2"). It is to be noted that it is proposed to time these operations so as to avoid coincidence with the population and agriculture censuses and thus even out the administrative load of the Bureau.

The current program in business and economic fields has been greatly developed in recent years, although it has been cut back considerably this year through reduced appropriations. Expansion of program has occurred particularly in the field of current statistics on manufacturing and in distribution. In the field of manufacturing, current statistics on production and shipments and other information is made available for important commodities on a monthly, quarterly, or annual basis, depending on the importance and volatility of the industrial activity involved. In the field of distribution, current statistics are made available on changes in the volume of sales or receipts in retail service and wholesale trades together with statistics on inventories in wholesale trade on a monthly basis and retail trade annually.

Monthly statistics are made available on the import and export trade of the United States including quantities and dollar values, together with supplemental information on vessel and air shipments, vessel entrances and clearances, and gold and silver movements. The foreign trade data are made available by commodities and by countries of destination and origin.

Current information is also made available on revenues, expenditures, and employment for state and local governments.

Although the Census Bureau is the largest general statistical agency in the government, its program is but a portion of the total Federal Statistical program which is a greatly decentralized program. The statistical output of the Bureau of the Census both supplements, and is supplemented by, the programs of other agencies. To obtain a comprehensive view, therefore, of the statistics available in any given field, it is necessary to study the program of other agencies as well as that of the Census Bureau.[1]

METHODOLOGY

Much of the substantive program which is in effect or which is under consideration is possible because of the methodological advances based on research in the Bureau of the Census and in the statistical field in general.

Among the more important of the methodological improvements to which the Census Bureau has made important contributions are the developments in the field of sampling. Important innovations both in theory and in practice have made it possible to design highly efficient samples not only for human population but also for business and industrial establishments. Efficient use of sampling methods have resulted in great reductions in costs; in increased speed in collecting, tabulating, and publishing results; and in improved quality of data because of the way in which experimental and research work has been facilitated.

Another field of importance in producing speedy, accurate, and low-cost statistics is machine tabulation. The punch card method of tabulation had its origin in the Bureau of the Census which still designs and produces much of its own tabulation equipment. Continuous search for improvement in tabulation equipment and technique has produced great dividends in the reduction of cost and increasing the speed and accuracy of tabulation. At the present time, among other developments, an electronic computing and tabulating machine is under construction which may completely revolutionize census tabulation techniques.

Important developments have also occurred in the Census Bureau's cartographic work. Good maps are essential for successful census undertakings. Among the more important recent developments in this area of the Bureau's work is the utilization of aerial photographs and detailed maps for urban areas showing individual structures on city blocks for sample selection and enumeration control.

Another development worthy of at least mention in passing is the utilization of methods of offset reproduction for the publication of the census reports. The use of these cheap quick methods of reproduction has made it possible greatly to increase the volume of statistics published, without increasing costs.

---

[1] The recent volume *Government Statistics for Business Use,* edited by P. M. Hauser and W. R. Leonard (John Wiley & Sons, New York, 1946) presents a comprehensive picture of many of the major types of statistics available from the Federal Government.

The statistical output of the Bureau of the Census is so large and varied that it is important that there be maintained a systematic catalogue of publications. A Catalogue and Subject Guide is available for the year 1946, and will be issued annually hereafter. In addition, a monthly list of publications is also issued. A study of these catalogues is the first logical step in getting acquainted with the statistical program of the United States Bureau of the Census.

Résumé

Le programme, les méthodes et la gestion du Bureau de Recensement ont été soumis à un examen au cours des dernières années dans le but:

1. d'augmenter l'utilité des résultats pour les usagers des chiffres;
2. d'améliorer et moderniser les méthodes et techniques;
3. d'augmenter l'efficience administrative. Bien que d'importants changements dans le programme, la méthodologie et l'organisation aient été apportés ces dernières années, beaucoup d'entre eux ayant été stimulés ou rendus nécessaires par les circonstances de guerre, la tâche de projeter des plans utiles, sûrs, équilibrés et efficients, est un procès continu, dont on s'occupe encore activement.

Les activités statistiques du Bureau de Recensement embrassent six domaines distincts: population (y compris main d'oeuvre et habitations), agriculture, industrie, distribution (y compris le commerce de gros, de détail et services personnels), administration publique, et commerce extérieur. Pour les cinq premiers domaines, le programe peut être divisé en deux phases distinctes: (1) la réunion et la publication de données de recensement, (2) la réunion et la publication d'informations courantes. En matière de commerce extérieur, des statistiques mensuelles sont réunies et distribuées d'une manière continue. Les statistiques courantes occupent de plus en plus la première place, de sorte qu'une grande importance doit être attachée à la réunion et présentation rapide et claire des données.

Programme de travail.

Présentement, un recensement complet de la population a lieu tous les dix ans dans les années dont le dernier chiffre du millésime est un 0, et un recensement complet de l'agriculture a lieu tous les cinq ans dans les années dont le millésime se termine par un 0 ou un 5. Un relevé par sondage de la population des Etats-Unis est fait mensuellement afin de procurer, pour l'ensemble du pays, des renseignements actuels sur la main d'oeuvre et sur des caractéristiques générales de la population. Ce relevé est basé sur un échantillon restreint de 25,000 ménages. Des plans tendant à améliorer l'utilité de statistiques démographiques et agricoles comportent la possibilité d'un recensement quinquennal de la population opéré simultanément avec le recensement quinquennal de l'agriculture. On considère aussi la possibilité d'un recensement annuel de la population et de l'agriculture par la méthode de sondage.

Dans le domaine économique et des affaires, on opère, pour le moment, un recensement décennal de la distribution pour les années dont le millésime se termine par un 9; un recensement de l'industrie minière pour les années dont le millésime se termine par un 9; et un recensement de l'administration des Etats et des administrations locales pour les années dont le millésime se termine par un 2. On est autorisé à effecteur un recensement de l'industrie manufacturière tous les deux ans; il se rapporte à l'activité industrielle dans les années impaires. Une nouvelle législation est en préparation qui autorisera un recensement quinquennal de la distribution, de l'industrie manufacturière et des industries minières, et aussi un nouveau recensement du transport, excepté les chemins de fer (pour les années dont le millésime se termine par un 7 et un 2). Il est à remarquer que l'on se propose de fixer l'époque de ces recensements de manière à éviter qu'ils coincident avec le recensement de la population et le recensement agricole, et, par conséquent, de répartir d'une manière uniforme le niveau d'activité.

L'extension du programme de statistiques courantes économiques et des affaires a eu lieu particulièrement dans le domaine de l'industrie manufacturière et de la distribution. Des statistiques mensuelles sont établies sur l'importation et l'exportation aux Etats-Unis comprenant les quantités et la valeur, en même temps que des renseignements supplémentaires sur les transports par mer et par air, les entrées et sorties de navires, les mouvements de l'or et de l'argent. Des données courantes sont également rassemblées sur les recettes, les dépenses et le personnel des gouvernements locaux et des Etats.

Méthodologie.

Parmi les améliorations méthodologiques importantes auxquelles le Bureau de Recensement a contribué dans une large mesure, il faut signaler les développements dans le domaine de l'échantillonnage. Un emploi efficient de méthodes d'échantillonnage a permis une forte réduction des frais, une augmentation de la rapidité dans la réunion, la tabulation et la publication des résultats; et une amélioration de la qualité des données par suite de la facilitation du travail expérimental et de recherche.

Un autre domaine important pour l'obtention de statistiques rapides, sûres et à peu de frais, est la tabulation mécanique. La méthode de tabulation au moyen de cartes perforées a son origine au Bureau de Recensement. Une machine électronique calculatrice et tabulatrice, actuellement en construction, pourra causer une véritable révolution dans la technique de la tabulation de données de recensements.

D'importants développements ont été également réalisés dans le travail cartographique du Bureau de Recensement; l'emploi de photographies aériennes et de cartes détaillées de territoires urbains montrant la structure de blocs, aux fins de la sélection d'échantillons et le contrôle d'énumérations.

La production statistique du Bureau de Recensement est si large et si variée qu'il est important de maintenir un catalogue systématique de ses publications. Un catalogue et index des matières est disponible pour l'année 1946; il sera publié annuellement par la suite. En outre, une liste de publications paraît mensuellement.

THE PROGRAM OF THE U. S. BUREAU OF LABOR STATISTICS

by Ewan Clague
Commissioner of Labor Statistics
(United States)

The United States Bureau of Labor Statistics is one of the principal agencies in the Federal Government for the collection and analysis of current economic statistics. It is one of our oldest continuing statistical organizations, and, through the 60 years of its experience, has gradually been given responsibility for a broader field of work than is suggested by its name. Its work is not restricted to labor statistics in a limited sense but, rather, covers what might generally be described as industrial and economic statistics of importance to the welfare of wage earners and of industry. Its charter, in American law, is a broad one:

> "...to acquire and diffuse among the people of the United States useful information on subjects connected with labor, in the most general and comprehensive sense of that word, and especially upon its relation to capital, the hours of labor, the earnings of laboring men and women, and the means of promoting their material, social, intellectual, and moral prosperity."[1]

Under this broad charter its program has been constantly changing over the 60 years of its existence to keep pace with changing times. Today its work is divided into two important sections—the publication of monthly or quarterly economic statistics on a prompt and current basis for such important aspects of the American economy as industrial employment and pay rolls, prices and cost of living, earnings of American wage earners, industrial accidents, and the number of work stoppages. In addition, it covers a wide field of economic analysis growing out of these and other data in order to present the results of fact-finding surveys bearing upon policy questions in the field of labor.

One important fact must be kept in mind in considering the statistical program of the Bureau of Labor Statistics: it has no compulsory powers to obtain reports. It has no regulatory powers; it enforces no law. All of the basic data that it collects are provided on a voluntary basis. The Bureau must therefore frame its program in such a fashion that the importance of its work as a fact-gathering agency will commend itself to industry and to labor so that they will supply the basic data upon which its statistics and its economic reports are based. The data reported by industry, labor and individuals to the Bureau are kept confidential and are never divulged without written consent of the original reporter. We in the Bureau regard this voluntary cooperative reporting structure as a great asset in a democracy. It has enabled the Bureau to get accurate statistics for information purposes during the war at a time when some regulatory agencies found it difficult to do so because of fear on the part of the reporters that they, as individuals, might be penalized for some infraction of the law, often unwitting.

---

[1] June 13, 1888, c. 389, sec. 1, 25 Stat. 182; Mar. 4, 1913, c. 141, sec. 3, 37 Stat. 737.

For this reason the collection procedures of the Bureau of Labor Statistics which I wish to discuss today may not have direct applicability for governmental agencies in countries where fact-gathering functions may rest upon authority to regulate or to compel reports. Our procedures may be more directly applicable to those used by private research organizations elsewhere. Nevertheless, I hope that some of the lessons that we have learned in our 60 years of experience will be of value to you even though the nature of our powers and our problems is different from yours. After all, statistical methods are universal to a high degree.

In discussing the Bureau's program today I want to describe in some detail both the substantive kinds of work that are undertaken and the technical phases of the program which are perhaps of greater concern to you—that is, the purely scientific or statistical procedures such as those involved in collection of data, sampling, the analysis of variance, and index number construction. The latter, unfortunately, are perhaps less publicized and less well-known than are the substantive aspects of our work. Accordingly, after a brief review of the Bureau's history, I shall consider first the substantive and later the technical aspects of the statistical programs upon which we have been working.

HISTORICAL BACKGROUND OF THE BUREAU OF LABOR STATISTICS

In response to increasingly wide demand in the United States for information of interest concerning labor, the "Bureau of Labor" was established in 1885 as part of the work of the Department of the Interior. When on March 4, 1913, a Department of Labor with executive rank was created, the Bureau of Labor Statistics, already 28 years old, was established as a part of the new Department. From 1913 to the present the Bureau has remained with the same basic functions as a part of the Department of Labor.

From the very beginning, the major projects of this Bureau have revealed both the wide variety and the comprehensive nature of the interest in information on labor and industrial development in the United States in the latter part of the 19th century. Its program in those early days was guided by the intent of Congress that it should be comprehensive, and no task in the economic or sociological realms was too large to undertake; nor did there appear to be any areas of interest to labor which were barred from study by this new research organization.

The first ten annual reports, issued between 1886 and 1895, dealt with the following subjects: Industrial Depressions; Convict Labor; Strikes and Lockouts; Working Women in large Cities; Railroad Labor; Cost of Production: Iron, Steel, Coal, Textiles, and Glass; Industrial Education; and Building and Loan Associations. Various special reports which were made in this same period dealt with marriage and divorce, labor laws in various States in the United States, and a variety of labor problems and methods of handling them in other major countries of the world. The first large-scale comprehensive study of wage-earner expenditures was made as early as 1901, and even in terms of modern techniques it was excellently conceived and carried out.[2] While some of these functions have long since been assumed by other agencies, most of them are still principally a concern of this Bureau.

In addition to these annual and special reports, under the direction of Carroll D. Wright, the Commissioner, the "Bureau of Labor" in those early years served as an important statistical and research arm for the United States Congress in conducting such special investigations as that for the Senate Committee on Finance on retail prices and wages. This type of function still continues.[3]

---

[2] *Eighteenth Annual Report of the Commissioner of Labor,* "Cost of Living and Retail Prices of Food" 1903, Washington; G.P.O. 1940; Department of Commerce and Labor Document No. 23, "Bureau of Labor."

[3] The description of the need for this research and the principal cause that led to the adoption of the Senate resolution might have been written in 1947. It was:

> "..undoubtedly the existence of a constant demand from legislators and economic students for reliable statistics in regard to the course of prices and wages in the United States. The absence of such statistics has led to a constant and indeterminable dispute over facts in every economic discussion. Without them it has been impossible to judge even with approximate accuracy the progress of the people of the country and the changes which have taken place from time to time in their condition.
>
> "There was no data in existence by which the actual or relative status of wage earners could at any time be accurately measured. As no adequate attempt had been made to cover this vast field for inquiry, the committee

The extension of this same kind of research on prices and wages as reported by the Senate Committee was concluded in its report on "Wholesale Prices, Wages, and Transportation" in 1893.[4]

One of the first major studies was a report on strikes and lockouts in American industry between 1881 and 1886, presenting details by states, industry, and causes of strikes, and amount of wages lost by the workers involved in the strikes. From this almost monumental first task the industrial relations work of the Bureau of Labor Statistics has grown. Simultaneously with the developing complexity of industry, working agreements have become increasingly complicated and require much more comprehensive study and analysis.

Over 50 years ago the first simple studies of productivity in American industry were started by the Bureau. Rapidly changing technology in the production of many goods, it was recognized, had profoundly affected the demands from labor and dollar distribution among the various industries and different occupations in the United States. It had even then long been realized that increasing productivity was the key to the higher wages paid in the United States and to the high American standard of living. These studies on productivity, both in their substantive and technological phases, still represent an important part of the Bureau's program.

With the passage of time it became clear that certain basic statistics with respect to major economic facts must be regularised and that

---

> were impressed with the necessity of exercising great care in the conduct of the important investigation committed to their charge."

The inquiry for this period was directed to the ascertainment—

> "First. Of the course of retail prices of two hundred and fifteen articles of general consumption in seventy cities and towns in different parts of the country.
>
> "Second. The course of wholesale prices for the same articles at seven principal distributing centers.
>
> "Third. The farm prices of leading agricultural products at representative points selected by the Department of Agriculture.
>
> "Fourth. The relative rates of wages in the cities and towns where price statistics were collected in fifteen general occupations.
>
> "Fifth. The course of wages in special industries.
>
> "Sixth. In addition to the inquiries as to the course of prices and wages in the United States, the Committee, through the agents of the Department of Labor, undertook to collect similar statistics for retail prices and wages in Great Britain covering the same period."

("Retail Prices and Wages," Senate Report No. 986, 1st Session, 52nd Congress, Part 1, pp. i and ii.)

[4] Senate Report No. 1394, Finance Committee, 2nd Session, 52nd Congress.

continuous series of statistical information must be developed and regularly published. The realization of the various statistical series that were important in appraising the condition of labor came only over many years. For instance, the first series on industrial employment assembled by regular reports from business firms were started in 1923; the wholesale price series were begun in 1902 (and were subsequently carried back to 1749); and the first cost of living index was published in 1921 and later carried back to 1913. Some of these series were developed because of recognized needs during periods of peacetime in the United States. Great impetus to many of them, however, was given by the statistical needs of World War I, the long depression of the 1930's, and of World War II.

Review of this early history of the Bureau shows that the program of the Bureau of Labor Statistics was one which basically dealt with the development of statistical series as need for them became apparent and with specific investigations designed to answer important economic and social questions of the day, as they affected labor. It is in reflection of such historical background that the current substantive and technical program of the Bureau has meaning.

Current Substantive Program of the Bureau of Labor Statistics

Of first priority today, though not necessarily of the greatest importance, is the maintenance of the many recurrent statistical series which have been developed by the Bureau. I shall not take time here to describe them in any detail or to discuss the stimulating statistical problems which they involve. I have asked that copies of the *Monthly Labor Review* be made available to all who would like them, so that you can review the principal statistics which are published month after month in its "Current Labor Statistics" section. Although all of the statistical series that the Bureau produces are not printed in this section, the major ones are there. They have been classified for convenience into the following 6 major groups:

1. *Employment and Pay Rolls*, reported voluntarily every month by about 120,000 business establishments in all principal lines of trade and industry, in cooperation with state agencies.

2. Related data on *labor turnover* in manufacturing industries.

3. *Hours worked and weekly and hourly earnings* of factory and other workers.

4. *Prices and Cost of Living*, including the official indexes of wholesale prices, weekly and monthly, and the monthly index of consumers' prices for 34 large cities and the U. S. average.

5. *Work stoppages* in American industry and the man-days of idleness they cause.

6. *Building and construction* of various types, undertaken each month by private and public agencies and individuals with particular emphasis on home building.

Other important series, issued less frequently, are on productivity in certain manufacturing industries, industrial accidents, and wage rates.

In addition to the development and maintenance of these statistical series, the Bureau continues to have responsibility for a wide variety of comprehensive investigations dealing with many phases of American labor and industry, and of international labor problems as they impinge upon domestic issues.

These include periodic comprehensive studies of wage rates and wage structures and of price structures; the outlook for employment opportunities in various occupations and industries; a new series of reports on man-hours required to turn out certain representative products of American industry; studies of the effects of long working hours; and of the production of impaired workers in industry. Most of these studies are conducted solely by the Bureau of Labor Statistics, but, in addition, it continues to act as a special statistical service agency for Congress and for other governmental and private agencies. Just before the war, for instance, the Temporary National Economic Committee —a joint legislative and executive body—asked the Bureau to make a number of studies dealing especially with problems of prices.[5] Some of these were landmarks of research in their fields. Throughout the war special studies were conducted for the War Labor Board, the Army and Navy, the War Production Board,the Office of Price Administration, and the housing agencies, thus avoiding the necessity for organizing separate and duplicating statistical staffs in those agencies. As the war ended and the economy of the United States returned gradually to a peacetime basis, the nature of these special projects changed, as such a program must always change if it is to meet the needs of the times. Thus, for example, the Bureau prepared for the Advisory Committee of the Office of War Mobilization and Reconversion a comprehensive analysis of annual wage payment plans in order to provide information for unions and industries which were interested in guaranteed annual wages. A special study was also made of the experience of workers and their families who migrated to war production centers, lived through the period of high wages, and after the war were forced by circumstances to remigrate to peacetime production centers.

---

[5] "Price Behavior and Business Policy," Monograph 1; "Hourly Earnings of Employees in Large and Small Enterprises," Monograph 14; "Geographical Differentials in Prices of Building Materials," Monograph 33; from the Investigation of Concentration of Economic Power by the Temporary National Economic Committee.

At the request of the Congress, the Bureau in the past two years has been working on the development of a new wage earner's budget. In the past, family budget studies have been based on subjective standards. The Bureau's current approach, by contrast, is objective—it provides an econometric challenge and utilizes new methods of analysis which may be widely adopted in the solutions of a number of troublesome economic and sociological problems. Another major contribution growing out of a project which also embodies a new method of statistical handling of interindustry relationships is described in our report "Full Employment Patterns, 1950."[6] It presents an analysis of the operations of the American economy in quantitative terms in such a way as to facilitate investigation of the long-term economic problems involved in the maintenance or achievement of full employment, and systematic testing of alternative policies, and tracing the quantitative effects of given expenditures through the stages of the economic process.

Many of these studies have explored new fields of statistical methods and of subject matter. We hope to continue these explorations. In doing so, we recognize that contributions of the various other subject matter fields such as sociology, psychology, anthropology, public health and public administration must be utilized to a greater extent than in the past in understanding the social and economic problems around which our inquiries are often centered.

TECHNICAL PROGRAM OF THE BUREAU OF LABOR STATISTICS

So far I have dealt mainly with the substantive phases of the Bureau's program. Since this is a meeting of statisticians from a variety of substantive fields, I wish to spend the remainder of my time in describing some of our technical programs, in summarizing the general directions which they have taken. I hope that, later, you will have an opportunity to discuss some of these methods in greater detail with us at your leisure.

First, I should like to direct my attention to one of the most important and, lately, the least emphasized of problems in statistics—how to get sound data upon which statistical findings may be based. Since no statistical study can be any better than the accuracy and integrity of its basic data, our emphasis has been upon the continuous improvement and increased efficiency of all the original facts which we use. This requires the careful development of definitions of terms, a clear description of units of counting, In price collections, for example, it means the development of specifications giving detailed descriptions of the characteristics of the commodities being priced.

---

[6] Serial No. R.1868 of the Bureau of Labor Statistics, U. S. Department of Labor, reprinted from the *Monthly Labor Review*, February and March 1947.

We have come to depend very strongly on field collection by agents of the Bureau rather than by the use of questionnaires sent through the mails. For example, we have spent a great deal of time on the careful training of our wage agents who call at factories to obtain wage rates paid for various occupations in order that distinctions between occupations using various machine processes but often called by the same name will be clear to them. We have experimented in the use of local residents who are employed as part-time agents in order also to utilize the resources which exist in local knowledge of situations and acquaintance with local customs, and to save time and travel costs.

We have directed much attention to the relationships with our respondents, to make sure that those who are going to supply information will know what the problem is, and will understand why we are engaged in a particular inquiry. We find, in difficult and time-consuming studies, particularly new inquiries, that a personal call yields better understanding and better statistics. In our studies involving large groups of householders, we have depended, for a similar purpose, upon the radio, newspapers, and magazines to assist us in describing the importance of our requests and the need for accurate information.

Recently we have conducted some experiments with schedule design and have engaged in careful cross-checks of the accuracy of returns as we get them from mail schedules and from personal visits by our agents. We commend this type of check to your attention. We have lately, for instance, devoted considerable attention to the use of postcard schedules directed to householders for the collection of information on rents, alternating with personal visits by our agents. Another experiment has been with "split schedules" for obtaining information on one of our more elaborate inquiries, that on family expenditures. By "split schedules" I mean that instead of asking each respondent to provide information on all groups of family expenditures—a long and time-consuming task—we have drawn several samples and have asked one group to provide information on food purchases, for example, and another to provide data on clothing and housefurnishings purchases. All groups are asked to provide information on one common phase of their expenditures, in order to test the validity of the process.

One systematic test of the validity of basic data has been of interest to the public in the United States and may be of interest to you. It concerns the accuracy of data on food prices at retail. Information on retail prices of food for the Bureau's consumers' price index (formerly called the cost of living index) is collected by personal visits of agents to food stores. These agents are always identified as representatives of the Bureau of Labor Statistics. In addition to inquiring about prices, they usually have an opportunity of checking price tags. To test further the accuracy of their reports, the Bureau on several occasions sent other representatives into these same stores on the same day. These second representatives, as unidentified customers, were furnished with money to purchase food of similar quality. They reported the prices which they actually paid. (Lest you be concerned with the disposition of these purchases, they were turned over to Federal institutions for use that very day.) In several such tests, comparisons of matched prices indicated a small margin of error—never exceeding 2 percent for a typical bill of foodstuffs. This test had the merit of giving public acceptance to a series of figures that had been open to some question during periods when foods were scarce and prices were under wartime government control.[7]

The technical discussions of this Congress, in the main, have been related to problems of sampling. We are all happy that that can be the case, for we believe that in the last decade great strides have been made in our knowledge of samples, and in how they can be used and controlled. In view of the emphasis which has been given in other parts of the program to sampling, I shall not dwell on it here, except to say that the Bureau's statistical staff has participated in these developments, has modified practices in accordance with their own findings and those of others, and is making use of the best known sampling procedures at every point possible in our program. In fact, we have a number of interesting procedures in double and triple sampling in price collection. In collecting information on rents for family dwellings, various kinds of samples have been the basis for a number of tests of schedule responses while they were providing basic data necessary for rental changes in the consumers' price index. In employment statistics we must depend upon industry sampling except in the case of small retail and service firms which come and go from business and there we have recently made use of area sampling. In our surveys of family expenditures we have made some unique contributions, specially in regard to sample biases.[8]

An important part of our current technical program is the preparation of manuals describing step by step the statistical procedures which are used in computing our various indexes. I regret that it is impossible for us to provide you with printed copies at the present time. In the near future we expect to have our important procedures completely recorded in manuals. In the process of manualization, we are clarifying our procedures, and obtaining greater consistency in editing, coding, tabulation, and calculating, and we expect important advantages from this process of manualization as an operating statistical device.

---

[7] For the first of these tests, see "Appraisal of the U.S. Bureau of Labor Statistics Cost of Living Index: Appendix," reprinted from the *Journal of the American Statistical Association,* Vol. 39 (1944), pp. 57-95.

[8] "Family Spending and Saving in Wartime," Bulletin No. 822, Bureau of Labor Statistics, U.S. Department of Labor.

Since the Bureau of Labor Statistics is perhaps best known for its official index numbers of employment, pay rolls, and prices, I should report to you that we have made advances in our methods of calculating index numbers and, at the same time, have learned from experience that less emphasis needs to be placed on the more elaborate formulae for the construction of index numbers than upon the basic soundness of the data, to which I have already alluded. In our price indexes, for example, much progress has been made in careful definition of questions which the index is designed to answer, in careful analysis of the procedure used in substituting one price quotation for another, and in the rules for making allowances for goods that went out of the market during the war and subsequently returned to the market after the end of the war. The way in which these technical problems have been handled by the Bureau is described in the reports of two separate committees which made inquiry into the Bureau's "cost of living" index—one representing the American Statistical Association and the other the President's Committee on the Cost of Living.[9]

We have learned to center our attention upon the selection of data, upon representative coverage of the field which the index supposedly represents, and finally, upon simplicity in calculation. It is easier to explain a simple index number than an elaborate one. We have learned, too, that trends over time can be successfully measured with fewer items than we had in the past thought were necessary. For that reason we have moved in the direction of reducing the number of items used and hence the volume of calculations without in any way impairing the accuracy of the index. We still assemble price information on many other items, because of their value for analytical purposes.

One statistical technique that has long been important in the United States is the development of estimates of total volumes from partial data—partial either because they are based upon samples where the universe is difficult to test, or because there are gaps in data for certain important parts of the area to be covered. Important examples in the work of the Bureau are estimates of total nonagricultural employment and of the total volume of construction. Many similar situations must present themselves to you. In such cases, where it is essential to measure not trends alone but levels, there are usually two problems: how to get data to fill in gaps, and how to make sure that general levels shown by small samples are accurate.

---

[9] Both of these reports are contained in "Report of the President's Committee on the Cost of Living," published by the Office of Economic Stabilization. Available at U.S. Government Printing Office.

In the case of the volume of employment, which is shown by industry groups, both monthly index numbers representing employment changes and estimates of total number employed are required for all important nonagricultural lines of work. In the United States, unlike many of your countries, reports on employment and unemployment through unemployment compensation insurance are only partial because certain occupations and certain firms are not covered by the laws. Consequently, in developing totals of employment, monthly, from its limited group of reporting firms, the Bureau has made use of more complete reports from Censuses and from old age insurance and other records for certain types of industry and has periodically "adjusted" upward current figures to meet levels shown by these more complete reports and, using these comprehensive records, has rebuilt and improved its samples. In order to minimize the amount of adjustment, the Bureau has developed a technique for anticipating this downward bias. This technique reduces the amount of the adjustment which is normally required. To fill gaps in current reports from small firms, the Bureau has developed, through area sampling techniques, ways of estimating changes in employment which take account of increases and decreases in employment due to changes in the number of small business establishments in existence. In the United States, where many individuals open small stores or service establishments each year and, in turn, other such establishments go out of business, this component of employment becomes very important.

For construction, use has recently been made of area-type samples of counties in the United States to estimate the volume of construction going on outside the large cities, from which the only regularly reported data on volume of construction are available. Sample studies have also been made of the length of time required to complete various types of buildings in order to compile ratios by which it is possible to estimate the volume of construction actually under way at any one time from what we call "building permits," that is, official permits to begin construction.

I have mentioned these rather complex instances of estimates of totals from partial data knowing that these brief remarks can do no more than point to the problem, in order to give due importance to a type of statistical work to which we, in the Bureau, devote much attention.

It is often of very great significance for national policy here, as I am sure that it is in your countries. I hope that the Bureau's staff can explore with you, while you are here, some of the techniques that you have developed to meet similar problems, recognizing that each solution must depend in large measure upon the unique circumstances of the case.

Unfortunately, the technical findings of the Bureau of Labor Statistics, as well as those of other agencies, are not too well known. They have not been effectively and clearly presented to the public and to research workers. I intend that one of my major responsibilities in the next few years will be to see that the contributions of the Bureau of Labor Statistics in these technical fields are made available both to the statistical profession in the United States and to you in other countries.

Finally, I wish to go beyond the program of the Bureau of Labor Statistics, as an agency, and refer briefly to some of the more general statistical problems to which I hope statisticians in the United States will devote increasing attention. In solving them, I know that we can profit greatly by the experience of statisticians throughout the world. Some of these observations may seem elementary, but since the accuracy of statistics often depends greatly upon them, I believe that they are worthy of our most serious attention.

First, in the field of questionnaire design, the United States, along with some other nations in the last 15 years, has made great strides in certain central statistical controls which have been exerted largely through control of questionnaires, and much has been done to improve schedule forms. I fear that we have lagged, however, in applying statistical controls to test the effectiveness of the arrangement of questionnaires, either from the point of view of accuracy and balance of response, ease of response, or cost of data collection. I refer here to a variety of matters—the number of questions asked, the shape, size, color, and printing of the questionnaire, and other aspects of questionnaire construction. These are areas that need to be approached not only from a statistical point of view but with assistance from sociologists and psychologists, in addition to economists. Some approaches have been made to "deep interviewing" but they have so far only scratched the surface. In this Bureau we have also made some tentative experiments on recall responses, and it is our hope to conduct further tests on effectiveness of questionnaire design.

It is also my opinion that problems of tabulation and calculation have been given far too little direct attention by leaders in the profession. We need to ask our mathematicians to work out the most effective short cuts through whatever formulae they can develop in order that our calculation processes may be reduced to a minimum, giving the greatest possible accuracy and the least effort in checking. In large government agencies, such as our Bureau, these processes must be established on a mass production basis. I am sure that when we bend our best efforts to this particular task there will be an impact which has never been gauged on the kinds of calculating equipment with which our statistical offices in the United States are provided.

Closely related to tests of schedules and development of effective calculation processes is the better development of cost functions of our various procedures and products. In short, statisticians should turn their own powerful tools of analysis upon their work and its costs.

Another very important aspect of statistical administration for which the profession must assume increasing responsibility is the training of statistical workers. This is done not merely in the colleges, but on the job. Here, I believe that those of us who direct statistical bureaus must accept a large measure of responsibility. While we have need of a few broadly trained and highly expert mathematical statisticians, most of our work is in the simpler statistical techniques where experience in procedures and common sense and knowledge of the study in hand are of great importance. We can train by insisting that every statistician who undertakes a new project should describe his ultimate objective and his methods of reaching it in writing before he proceeds; by making sure that his work is carefully appraised and criticized and that he has an opportunity to benefit from the advice of the best technicians available. The American Statistical Association, in the past year, has set up a commission to study the training of statisticians. Meantime, I am sure that those of us in government agencies must use every training device which we can obtain, including the most effective teaching methods and the latest visual aids. This area of study is, of course, not peculiar to research bureaus dealing with statistics, but for statisticians it presents problems of a specialized nature to which agencies such as the Bureau of Labor Statistics can make an important contribution.

And last, I wish to suggest that we should apply statistical controls to the effectiveness of the presentation of our materials to the public —by press, radio, technical and other magazines, and other media; and to find out how clearly and effectively our results are, in fact, reaching the publics for which they are intended. It is a challenge to do so.

In summarizing the Bureau of Labor Statistics of the U. S. Department of Labor has had a program of statistical work from which there has been little deviation of purpose over the last 60 years and in which great developments in concepts and in techniques have been made. There has been increased realization of our responsibility as unbiased technicians and research workers providing tools for effective public decisions. It will be our endeavor to continue to extend our work into

new fields as new problems arise and into new technical areas as new tools are needed. In order to make our work in the Bureau of Labor Statistics more effective, we are assembling materials on techniques and methods that you are using in other countries. We welcome most heartily this opportunity to consult with you about them. I assure you that we will make every effort to place our data and our studies at your disposal and to conform to international standards for widely used statistical data, so that between your work and ours there will be an ever widening area of comparability and of understanding.

Résumé

Le Bureau des Statistiques du Travail des Etats-Unis est un des principaux offices fédéraux de réunion et d'analyse de statistiques économiques courantes. Ses activités se répartissent actuellement sur deux importantes sections: la prompte publication de statistiques concernant des aspects importants de l'économie américaine, tels que l'emploi et les salaires industriels, les prix et le coût de la vie, les gains des salariés, les accidents de travail et les conflits de travail. En outre, le Bureau s'occupe d'analyses économiques servant de base numérique à la politique du travail.

Il faut remarquer qu'il n'y a pas d'obligation légale de fournir des données au Bureau. Toutes les données sont obtenues par collaboration volontaire.

APERCU HISTORIQUE

Le "Bureau du Travail" a été établi en 1885 comme une division du Département de l'Intérieur. Lorsque le 4 mars 1913 un Département du Travail fut créé, le Bureau des Statistiques du Travail, comptant déjà 28 années d'existence, devint une division du nouveau Département, ce qu'il est encore toujours.

Les dix premiers rapports annuels, publiés entre 1886 et 1895, traitent les matières suivantes: crises industrielles, travail de prisonniers, grèves et lockouts; fer, acier, houille, textiles et verre; enseignement professionnel; associations de construction et de prêt. La première grande enquête sur les dépenses des familles de salariés a déjà été faite en 1901; les premières études sur la productivité du travail datent de plus de 50 ans.

A la longue, l'expérience a appris qu'il était nécessaire de régulariser certaines statistiques fondamentales et que des séries statistiques continues devaient être établies et publiées régulièrement. Par exemple, la première série concernant l'emploi dans l'industrie a été publiée à partir de 1923; les séries de prix de gros ont été commencées en 1902 (remontées, par la suite, jusqu'à 1749); le premier nombre indice du coût de la vie a été publié en 1921, et, plus tard, remonté jusqu'à 1913.

PROGRAMME DE TRAVAIL ACTUEL

La tâche principale, à présent, est d'assurer la régularité des nombreuses séries statistiques périodiques qui ont été développées par le Bureau. Elles ont été classifiées suivant les six groupes principaux ci-après:

1. Emploi et salaires,
2. Mutations de personnel dans les industries manufacturières,
3. Heures de travail et salaires hebdomadaires et horaires,
4. Prix et coût de la vie,
5. Conflits de travail dans l'industrie américaine et nombre de jours-ouvriers perdus qu'ils ont causé,
6. Bâtiment et construction.

En outre, le Bureau continue de s'occuper de nombreuses études d'ordre très varié se rapportant à plusieurs phases du marché du travail et de l'industrie américaine, et de problèmes internationaux du travail en tant qu'ils sont liés à des questions domestiques.

PROGRAMME TECHNIQUE

Un des problèmes les plus importants en statistique est le moyen d'obtenir des informations sûres sur lesquelles on puisse baser des résultats statistiques. Nous nous efforçons d'améliorer constamment et d'augmenter l'efficience des données de base utilisées. Les discussions techniques aux Conférences Internationales de Statistique ont concerné en premier lieu les problèmes d'échantillonnage. Au cours des dernières dix années, de grands progrès ont été réalisés dans notre connaissance dans ce domaine. Les statisticiens du Bureau des Statistiques du Travail ont contribué à ces développements et utilisent les meilleures méthodes d'échantillonnage connues.

Une partie importante de notre programme technique actuel est de préparer des manuels décrivant les procédures statistiques utilisées pour le calcul de nos divers indices. Dans un avenir prochain, nous espérons avoir une série complète de tels manuels. Malheureusement, les travaux techniques du Bureau, de même que ceux d'autres services, ne sont pas assez connus. Je considère comme un de mes principaux devoirs de mettre à la disposition des statisticiens, tant aux Etats-Unis que dans d'autres pays, les contributions méthodologiques du Bureau des Statistiques du Travail.

THE AGRICULTURAL STATISTICS PROGRAM

by Charles F. Sarle
Bureau of Agricultural Economics
(United States)

PERSPECTIVE

The more commercialized the agriculture of a country, the greater the demand for current statistics concerning all aspects of agriculture. Where agricultural products are grown only for farm and local consumption, there is little demand for agricultural statistics other than perhaps for an occasional agricultural census. The evolution of the demand for agricultural statistics has been well exemplified historically in the United States.

The publication of statistics on crops, livestock, and prices of agricultural products is the oldest activity in the Department of Agriculture, having been started in 1862 when the Department was organized. These statistics were begun after several decades of agitation by leading farmers for unbiased agricultural statistics concerning condition during the growing season, and harvested production of crops that were being produced commercially, especially cotton, wheat, and other grains. These farmers wanted to be protected, as they said, from the effect of false and misleading reports emanating from those who wished to buy their products at low prices.

As a nation's economy becomes more diversified and complex, demand for agricultural statistics increases. Demand is not only for broader coverage of agriculture, but also for other facts relating to the ever changing agricultural process, for statistics at more frequent intervals, and for greater accuracy. This demand is accelerated by the strain put upon a nation's economy by war. When national economies are subjected to world-wide depression and governments embark on production control and price-support programs, the demand for more and better agricultural statistics increases almost overnight.

KINDS OF AGRICULTURAL STATISTICS

At present, the demand for agricultural statistics, which the Department's program endeavors to meet with the resources available, covers a broad field. In addition to statistics concerning production, prices, marketing, farm employment, and wage rates, the Bureau of Agricultural Economics maintains such regularly recurring statistical series or estimates as 'Indices of farm production; farmers' cash receipts and expenditures; parity prices and parity income; gross and net farm income; estimates of per capita food consumption for individual commodities; changes in land values and volume of sales of farm real estate; farm taxes and and tax returns; interest paid and indebtedness; valuation of livestock and farm equipment; cost and returns for specific types of farming; estimates of farm population; migration to and from the farm; and similar series which are prepared by bringing together and studying large masses of data collected by this Bureau and other agencies.

To meet these needs, the Bureau of Agricultural Economics published numerous basic statistical series throughout the year. Included are series of current estimates of production, stocks and prices of over 150 farm products, and also farm wages and employment. (See Appendix A—"Agricultural Statistical Series Published by the Bureau of Agricultural Economics")

ORGANIZATION AND COOPERATION

The Bureau of Agricultural Economics is general overseer of the economic research and statistics program of the Department of Agriculture. Six divisions of the Bureau are engaged primarily in gathering crop and livestock statistics. These divisions are known collectively as Agricultural Estimates. There are six other divisions of the Bureau—the economic investigations divisions—which also construct and maintain regularly recurring statistical series, such as farm income and expenses, marketing margins, farm population, mortgage debt, land values, taxes, etc.

The Bureau maintains 41 State statistical offices which serve all States. The state agricultural statisticians in these offices are responsible for maintaining cooperative relations with State agencies and contact with farmers and other people for collecting statistical data relating to agriculture for their respective States; and for analyzing, summarizing and interpreting the information received from farmers and other sources which they forward to the Crop Reporting Board or to statistical divisions in Washington for review and assembly into State and national averages and totals.

Appendix B is a chart that shows diagrammatically the Bureau's nation-wide fact-finding organization, and how farm statistics are collected and distributed to the people who use them.

SOURCES OF DATA FOR CURRENT AGRICULTURAL STATISTICS

The data used in making agricultural estimates fall into five general categories:

(1) *Enumeration Data*: Data that are obtained by complete or nearly complete enumeration, such as the periodic agricultural censuses which have been taken every 5 years since 1920, and at 10-year intervals from 1840 to 1920; annual enumeration of crop acreages made by local tax assessors in about 12 States; and statistics covering the operations of various types of processors and handlers of commercial agricultural products, such as cotton ginnings, rice millings, peanut shellings, tobacco sales, sugar manufactured from sugarcane and sugar beets, vegetable processors, carlot shipments of fruits and vegetables, plant production of manufactured dairy products, shipments of livestock and grain, market receipts and quantities of certain agricultural products processed for sale. In some 31 States tax authorities report annually or biennially on the assessed numbers of live stock on farms.

(2) *Mailed Sample Data*: Questionnaires are distributed in two ways. Either they are mailed directly to lists of volunteer reporters who serve without pay or are distributed by rural mail carriers to about 10 percent of the farms along their routes. The rural-carrier surveys, conducted in cooperation with the Post Office Department, are made once a year (October) to obtain information on crop acreages harvested, and twice a year (June and December) to collect information on livestock numbers and production of livestock.

In addition, the 41 field offices maintain a number of special lists of volunteer reporters to whom various types of mailed inquiries are sent, such as prices received by farmers, prices paid by farmers, special commercial crops, milk producers, feeders of cattle, sheep, and lambs, acreage planted, acreage harvested, and production of crops, etc. A mailed sample also is used to obtain information from flour mills and grain elevators, cotton ginners, local slaughterers of livestock, creameries, fluid milk plants, hatcheries, and other processors of agricultural products.

The inquiries sent to the list of volunteer reporters include questions concerning crops, livestock, farm wages, etc. in their respective localities and also questions concerning acreage and production of crops and livestock on their individual farms. The rural-carrier surveys always relate to the individual farms of the respondents.

During the entire year a total of about of 9,500 separate schedules are mailed to lists of farmers and others including those distributed by the rural carriers. In all a total of approximately 9,800,000 schedules are distributed in this manner and about 2,600,000 or 26 percent are returned and tabulated.

The regular list of 80,000 volunteer farmer reporters receive a schedule the first of each month. Of this number about 24,000 schedules or 30 percent are returned and tabulated. The rural carriers distribute about 600,000 cards to farmers along their routes three times each year. About 160,000 cards or 27 percent are returned and tabulated. In June a mailed survey on crop acreages is sent to lists of about 344,000 volunteer reporters. Of this number about 75,000 schedules or 22 percent are returned and tabulated. In late October a mailed survey on crop acreages harvested, acres abandoned during the growing season and production are sent to about 320,000 volunteer reporters. About 88,000 schedules or 27 percent are returned and tabulated.

(3) *Field Observation and Contacts*: Information concerning crops and livestock is obtained by general field observation and personal contact with well-informed growers and other "key" people who buy or handle farm products. This approach is practically indispensable, for example, in making estimates of fruit and vegetable crops, the production of which in most areas is highly localized because of special requirements of soil, climate, and markets; in evaluating the damage done to crops by such disasters as hurricanes, droughts, freezes, floods, or hailstorms, and in estimating cattle and sheep on feed.

(4) *Objective Plant Counts and Measurements*: With certain important commercial and feed crops, objective plant counts and measurements are made in the fields in which these crops are growing. For example, just before harvest, yield per acre of small grains, corn, or soybeans can be estimated by taking small samples of specified areas from fields throughout the territory where the crop is grown. Objective methods are used in selecting the fields to be sampled and in selecting the small "sample areas" for harvest.

The percentage of infestation of boll weevil in cotton is measured by counting the sound and infested bolls. For citrus fruits "frame counts" are made in California and Florida in selected groves to determine the number of fruits on trees in comparison with previous years. Diameters of representative fruits also are measured to obtain an index of size. Estimates based on objective plant-count measurements are free from the bias of personal judgment of the observer or opinion of the reporter.

Another objective field measurement of crop acreages is obtained by a machine called the crop meter. This machine, which is attached to the speedometer cable of an automobile, is used to measure the frontage of each kind of crop along selected routes each year.

(5) *Preselected Interview Surveys*: In recent years sample interview surveys have been successfully used to a limited extent in making

estimates of certain kinds of agricultural phenomena, such as family and hired labour employed on farms and farm wage rates, farm income and expenses, inventories of livestock, crops, and farm machinary, etc.

Many kinds of important agricultural data are difficult or impossible to secure by mail and these must be obtained by enumeration or interview. The interview survey method makes it possible to reach all types of farms and farmers, including people who do not respond to mailed questionnaires. It is less expensive and time-consuming than a complete census.

In an interview survey, well-trained rural people are employed to interview all farmers (or others) in small areas of about five farms each selected beforehand on a representative basis. The chief thing is that a small but carefully selected cross-section sample of the population of farms or of farm operators is contacted and that the desired information is obtained by personal interview.

Statistics obtained in this way can be expanded into estimates for a small area within a State, for an entire State, for a group of States, or for the nation as a whole, depending upon how many farmers are interviewed and how great an area is covered.

If national totals only are desired, the number of farms in the sample may be quite small for certain items. But to get accurate estimates for regions, States, or areas within States, proportionately larger samples are necessary. Thus, if State estimates are desired, almost as many interviews may be necessary within each State as would be required throughout the entire country for corresponding national estimates. The accuracy of an estimated average or total depends on the number of observations included in the sample as well as on the frequency of occurrence and degree of variability of these observations.

The number of interviews necessary depends also on the type of information or estimates desired. For instance, it takes a comparatively small sample to provide dependable data for items or conditions common to a large proportion of farms—say, the number of farms raising livestock or producing corn. But a comparatively large sample must be surveyed to get accurate statistics about something common to only a small proportion of farms, such as specialized equipment like combines or corn pickers, or the number of hired farm laborers.

METHODS OF ESTIMATION

It is an oversimplification to say that the 5-year periodic census statisics are used as "bench marks" for current agricultural statistics and that the other data mentioned above are used only to provide a basis for estimates in intercensal years. Actually, the statistical series published by the Bureau for Agricultural Economics are in a large part estimates based on the relationships (regressions) that exist historically between mailed sample data and various presumably complete or nearly complete enumeration data including the Federal census, State assessors' censuses in those States where they are available, and/or various commercial check data mentioned under class 1 source of data.

One example of agricultural statistics entirely independent of census enumerations is agricultural prices. In fact, the price data collected monthly by the Bureau of Agricultural Economics during a census year are used by the Census Bureau in arriving at the value of agricultural products. The price statistics of the Bureau of Agricultural Economics—prices received and prices paid by farmers—are based almost entirely on mailed samples from special volunteer price reporters who are usually local dealers and merchants in farm communities throughout the country.

The Bureau of Agricultural Economics of the Department of Agriculture is responsible for the current series of statistics relating to agriculture. The series relating to agricultural production for the United States and for each of the 48 States—intentions to produce, prospective production, and estimates of production as well as prices received and paid by farmers—must be *timely to be of value.* For example, the monthly crop reports relating to conditions as of the first of the month are issued between the 8th and the 10th of the current month, the twice-a-year reports of hog production (The Pig Crop Report) are issued within 18 to 20 days of the end of the 6-month period covered by the report. Somewhat more time (about 6 weeks) is allowed for the December report of harvested acreage and production of all crops. The public has been used to this kind of a " timetable " since about 1865, or for more than 80 years. Any attempt to extend the time by even a few days would be highly unpopular. Consequently the planning of future developments and improved sampling techniques will need to keep this speedy timetable in mind.

The Bureau of the Census is responsible for the agricultural census which is taken every 5 years. The census statistics provide information in great geographic detail, that is, by counties, but there is necessarily a considerable time lag in the release of census statistics.

Statistics on General or Major Crops

The general crop statistics fall into two categories: (a) Estimates of acreage harvested, yield per acre and production; and (b) estimates of acreage planted, condition of the growing crop, and prospective production.

The Federal census and State assessors' censuses in many of the important North Central States, along with the annual (October) mailed rural carrier survey, provide a reasonably dependable basis for estimates of harvested acreages of the more important grain and feed crops. Cotton ginnings provide an annual check on cotton production. The mailed returns from volunteer crop respondents who report on average yields per acre for different crops in their locality, usually provide a dependable basis for State estimates for yield per acre for major crops.

The monthly condition of growing crops reported in "percent of a normal or full crop," when interpreted on the basis of average relationship (regression) for that month and crop between reported condition and final yield per acre, usually provides a reliable indication, as of that date, of the prospective out-turn per acre for most major crops.

The closer the date of the reported condition to the time when the crop has reached maturity, the higher the relationship between condition and final yield, and hence the more reliable is the indication of prospective yield per acre. The earlier in the growing season that condition is reported, the greater the possibility that subsequent weather can materially modify these early indications of yield per acre.

The reported condition of a growing crop fails to be a dependable indication of prospective out-turn with some crops in certain areas even late in the season. With some crops, the report of the "probable yield per acre" used on a relative basis from year to year is a better indication of harvested yield than the condition, as the latter tends to reflect only the vegetative appearance of the growing crop.

LIMITATIONS OF SOURCE DATA

Census statistics available every 5 years are most reliable on acreage, yield, and production of major general crops. However, even the census statistics on these crops are subject to some error because the owner or tenant-operator on the farm when the census is taken is not always the same person who operated that farm the previous crop season. If the census could be taken in the late fall, this source of error would be held to a minimum.

The census statistics of the production of fruit crops in census years, along with some annual check data on the commercial movement of fruit crops and the reported condition of these crops in percentage of normal or "full crop," provide a reasonably satisfactory basis for estimates of the production of the major fruit crops in most States.

Vegetable and Special Commercial Crop Statistics:

It is with vegetable statistics that census enumeration data are most incomplete and inaccurate. For example, the 1945 census schedule included only nine of the 29 vegetable crops for which the Bureau of Agricultural Economics makes official estimates. It has not been possible in the past to so select, train, and supervise census enumerators that they will consistently ask all the questions on the schedule relating to crops and other agricultural items that occur on only a small percentage of the farms in their respective enumeration districts.

It is extremely difficult also to obtain reliable information concerning these sporadic phenomena by means of the mailed questionnaire to lists of volunteer correspondents. When the production of a vegetable or other special crop is highly localized geographically because of exacting soil, climate, or market requirements, the most economical method of obtaining information concerning acreage and production is by personal contact with well-informed growers and other "key" people who buy or handle such crops. The other and more expensive alternative is either a complete enumeration of all the producers in each local producing area or an interview sample of a large proportion of these growers.

Livestock Statistics:

If the census could be taken at the same time of year each time and taken during a short period, the census livestock statistics of numbers of different kinds of livestock on farms would provide a much more satisfactory basis for annual inventory statistics. Census statistics of the production of livestock and livestock products during the preceding year are in considerable error (memory bias) because of the seasonal nature of such production. The same difficulty exists with any agricultural phenomena with high within-season variation, such as farm family and hired labor, farm expenses and income, etc.

Census data on milk production covering an entire calendar year are subject to considerable bias as they tend to be influenced by the level of production at the time of year when the census is taken. For example, if the census is taken in January—a period of low production in most areas—milk production for the previous year will be greatly understated to the enumerator. A measure of the seasonal pattern of milk production per cow is obtained from the monthly reports of regular crop reporters and from the June and December returns from the rural mail carriers.

Statistics of manufactured dairy products include a complete enumeration of the plant output of the major products, together with current estimates on a monthly or weekly basis for the more important ones. Each year a complete enumeration of factory output of about 45 different kinds of dairy products is obtained. Dairy plants keep comparatively accurate records of production and in many States collection of data is assisted by State laws requiring the firms to report the amounts manufactured. In 28 of these States the Bureau has entered into cooperative agreements with certain State agencies to provide for the joint collection of these reports. These facts, plus a diligent follow-up program in obtaining reports for delinquent plants, have resulted in statistics that closely approach 100 percent completeness on all but a few minor items.

Poultry and egg production statistics also have high intraseasonal variability. Consequently, the census taken at only one time during the year does not provide reliable data on numbers of chickens produced or sold or eggs produced. However, the mailed sample taken monthly does provide a useful measure of the seasonal pattern of poultry and egg production.

Farm Employment and Wage Rates:

The Bureau of Agricultural Economics series on farm employment, family and hired labor (monthly), and farm wage rates (quarterly) are based on a mailed inquiry to the regular volunteer reporters of the Bureau. The sample returns on farm employment are adjusted during the season by factors derived from a comprehensive study of census data and farm labor requirements obtained from farm-management investigations made in different parts of the country. During the last few years; five nation-wide interview and several special mailed surveys have been made on farm employment and wages. It was found that on the national level there was very little difference between the wage rates obtained by the interview sample and the mailed sample.

An analysis is now being made of data on farm employment and wages obtained by the two methods of sampling for the purpose of determining more satisfactory within-season adjustment factors for use with the mailed sample, pending the time when resources will permit a quarterly interview survey of farm employment. Farm employment varies greatly with the seasons of the year, and the farms of crop reporters who report monthly are not sufficiently representative to provide a sample that can be used without seasonal adjustments.

MAILED VERSUS PRESELECTED INTERVIEW SAMPLING

The mailed sample has several outstanding advantages and some serious shortcomings as compared with the preselected interview sample of a representative cross-section of farms. The mailed sample is inexpensive both in time and money, and the returns can be obtained within a few days through the 41 field offices of the Bureau.

When the results of mailed samples are used on a regression basis with a reliable series of census, assessors, and/or commercial check data, useful State and national estimates can be made for major crops and classes of livestock in most States.

On the other hand, the mailed sample has serious limitations. As it is not a representative cross-section sample, only statistical averages and ratios can be used from it. These, in turn, must ordinarily be used on a regression or relative basis in making estimates. However, mailed samples on prices are an exception to this generalizatiotn. Other highly useful statistics, such as frequency distributions, frequencies of occurrence, and totals cannot be derived satisfactorily from the mailed sample.

Selectivity of Mailed Samples:

In using the mailed sample it is recognized that it is selective and not fully representative of the population for which it is drawn. It is selective at two levels:

(1) The usual mailing list to whom mailed questionnaires are ordinarily sent is not a representative cross-section of the population of farm operators, unless it has been recently drawn from a complete census or is a list of operators obtained by a preselected interview sample survey.

(2) The response obtained by mail is never 100 percent complete, hence there may be selectivity in the responses that are obtained. The differential response of a mailed sample may vary greatly because of the effect of many unknown psychological, social, and economic factors over which the statistician has little or no control. The differential response is especially serious in sampling agricultural phenomena that occur on a relatively small proportion of farms. Experience has shown that a mailed questionnaire designed to obtain such information has better "drawing power" when a number of questions are added that relate to other phenomena that occur on practically all farms.

The differential response to a mailed inquiry is not necessarily constant from one season of the year to another, or from season to season, or from area to area, as these little understood psychological, social, and economic factors change in time. For example, a different response would be expected on a cotton acreage inquiry when the price of cotton has dropped sharply than when the price has remained fairly constant over a period of a year or two. Part of the changing selectivity of the returns from the mailed sample is caused by varying proportions of "in and out" producers of an agricultural product from season to season.

Considerable research and experimenting needs to be done with mailed sampling when a true cross-section sample of farm operators is used as a list to whom the mailed inquiries are sent. Using such a list eliminates the first cause or level of selectivity with mailed samples as they are now generally used by the Bureau. Such lists could be drawn from the tax assessors' annual census records in the States where they are available.

The respondents of an annual preselected interview sample also would furnish such a list.

The research should be directed toward the discovery of psychological, social, and economic factors that give rise to the differential response to mailed inquiries of different kinds. To the extent that these factors can be determined, "control items" could be included in the original interview schedule for use in stratifying and weighting the returns of subsequent mailed samples, thereby minimizing selectivity at the second level mentioned above.

The effect of repeated "follow-up" of the same schedule to the persons who fail to respond on the first request needs further investigation and may lead to the development of a basic theory of nonresponse that can be expressed mathematically and ultimately used in estimating for the farms of nonrespondents. The Bureau statisticians have other ideas along these lines that should be tested. Every effort should be put forth to combine the advantages of interview and mailed samples in order to maximize accuracy and speed in getting sample data at minimum cost.

FUTURE NEEDS FOR AGRICULTURAL STATISTICS

The basic long-run objectives of the agricultural statistics program of the Bureau of Agricultural Economics are threefold. They may be summarized as follows:

(1) *Statistical:* To improve the accuracy of the official agricultural statistics of the Bureau and to develop additional essential statistics concerning agriculture that are needed in our complex industrialized national economy.

(2) *Research:* To provide agricultural research economists and rural sociologists with representative observational data from a true cross-section sample of farms, farm operators, and farm people, or of specified subpopulations thereof, which would serve as a valid basis for testing provisional and alternative economic and social hypotheses concerning the behavior of agricultural phenomena, including farm people themselves.

(3) *Federal-State Cooperation:* To develop in each State a sound integrated cooperative sampling program for providing State interests with agricultural information needed at the State level, and to provide State economists and rural sociologists with representative observational data for research purposes.

The basic requirements of an agricultural statistics program to meet the above objectives include:

(1) *One large annual nation-wide interview sample:* This sample should be of sufficient size to provide reliable State estimates for about

three-fourths of the States—the larger States in terms of number of farms—and for groups of smaller States. The State estimates from this sample would be limited primarily to agricultural items that occur on about 25 percent or more of the farms in a State or in a group of small States. However, national and regional estimates could be made for agricultural items that occur on a smaller proportion of farms. The minimum size of the annual nation-wide sample to meet these objectives would be about 60,000 farms, with the sample distributed in about 2,400 counties. Eventually this sample should be increased to 100,000—120,000 farms distributed in all agricultural counties.

It would be necessary to establish a current list of the large commercial farms and use a higher sampling rate with these.

West of the hundredth meridian (in the far western and Pacific Coast States) it would be necessary to establish and maintain currently a list of the larger cattle and sheep growers who use public grazing lands for their herds and flocks, and to delineate the irrigated and dry-land farming areas on appropriate maps in order that the sample segments may be drawn predominately from these strictly agricultural areas. Delineation of nonfarming, cut-over, and forest lands would also make it possible to increase materially the efficiency of area sampling in many other States, such as Michigan, Wisconsin, and Minnesota, and some of the Appalachian, Atlantic Coast, and Gulf Coast States.

(2) *Three quarterly interview surveys a year:* If more accurate statistics of agricultural phenomena that have a *high within-season variability* are to be obtained, interview surveys must be made oftener than once a year. These phenomena with high intraseasonal variation include family and hired labor, wages, production of milk, eggs, and livestock, the incidence of accidents and sickness to the farm population, and farm income and expenses. The frequency of occurrence of many of these items during a 3-month period is usually relatively low. A longer period than 3 months is too long for detailed and accurate memory by most farmers and would result in considerable "memory bias."

These quarterly surveys should probably be made in January, April, and July.

(3) *Special mailed samples to the respondents of the annual interview sample:* Special mailed samples to such a cross-section list of respondents would provide a much better basis of sampling than is now in use by the Bureau in making a number of its more important agricultural surveys each year. For example, if the annual interview sample were made in October or November, a mailed sample to these respondents could be made:

(a) *In June* each year, on hogs and milk cows, and on crop acreages planted, eventually to replace the June rural-carrier livestock survey and the June acreage survey to voluntary correspondents. These two surveys are among the more important ones made during the entire season.

(b) *In January* on numbers and kinds of livestock on farms on January 1 as a basis for the annual estimates of livestock inventories. The October inventories would be brought up to January 1.

(c) *In April* to the respondents of the annual survey in the *12 North Central States* to serve as a basis for a preliminary estimate of the size of the spring pig crop. There is an insistent demand on the part of the livestock industry for such statistics.

(d) *In December* on hogs and the fall pig crop, eventually to replace the December rural-carrier livestock survey.

(e) *In March* on acreage planting intentions. The returns would be compared directly with the crop acres harvested the previous fall which are obtained on the annual interview survey.

(4) *"Flash" surveys:* There is keen interest in a sampling scheme that would get *quick* results from a small sample on what farmers are thinking and planning to do when they are subjected to the impact of forces (economic, weather, and economic-political) which are beyond their control. It is likely that surveys of this kind might be directed to specific segments of the farm population, such as cattle or lamb feeders, commercial hog or dairy products, or producers of special commercial crops, such as citrus or potatoes. For example, such surveys might be used to discover farmers' reactions to a proposed control-price support program with a specific agricultural commodity.

A large annual interview survey of 120,000 farm operators would provide an excellent basis for small but representative subsamples of operators falling into these special categories. Representatives of the Bureau's 41 field offices, with some help in the form of interviewers, could handle one of these small surveys of 1,000 to 3,000 farms rather quickly.

(5) *Vegetable and fruit crops statistics:* The statistics that have the poorest foundation of census and other enumeration data and some of the most difficult problems of sampling and estimation are those relating to vegetable, fruit, and specialty crops. If this problem is to be solved, it will be necessary to *establish and maintain current lists* of commercial producers of vegetables, fruit, and specialty crops, State by State, which could then be sampled by mail and interview methods either singly or in combination, to provide a sound statistical basis for the statistics of acreage, production, and marketings of such crops.

In the case of tree fruits, complete enumerations are needed every 6 or 8 years in all chief commercial producing areas to obtain information concerning the *kind, age, variety, and production of tree fruits.* Such a census is necessary as a starting point for subsequent sampling and subsampling.

(6) *Special small nation-wide interview economic surveys:* If fully reliable statistical series are to be developed and maintained on a national and regional basis for such economic items as gross and net farm income, farm expenditures, nonfarm family income, indebtedness, savings, etc., it will be necessary to make special small nation-wide interview surveys, utilizing highly qualified interviewers who have had adequate training in the application of sound, practical social-psychological principles to the interview situation. These interviewers must be carefully supervised. In fact, considerable research and experimentation is needed to determine the best methods for minimizing what might be called "income tax bias" that is so prevalent in the reports of farm income. It will be some time in the future before these methods are sufficiently developed to justify a nation-wide sample of sufficient size to provide useful State estimates on farm income.

Conclusion: It is our conviction that a successful program of interview sampling must have three essential attributes. These attributes are like the three legs of a stool or table. Each must be strong and dependable or the stool is useless. These three attributes are:

(a) Statistically sound and efficient sample design

(b) Psychologically sound and practicable schedule design and interviewing methods

(c) Practicable and efficient organization and operation in the Washington office and the 41 field offices of the Bureau, emphasizing particularly the most effective practices of selecting, training, and supervising interviewers.

APPENDIX A

AGRICULTURAL STATISTICAL SERIES PUBLISHED BY THE BUREAU OF AGRICULTURAL ECONOMICS

As the farmers' principal economic fact-finding and research agency, the Bureau of Agricultural Economics publishes many statistical series about farm products, farm people, and many phases of farm business.

Primary Statistics:

The types of agricultural statistical series that are obtained by sampling are summarized below. All of these series are for the United States (excluding possessions) and most of them also are by States.

(1) Crop acreages (108 crops)

(2) Conditions of crops and pasture; prospective and harvested yields per acre and production (136 crops)

(3) Stocks of products on and off farms
(4) Numbers of livestock and poultry on farms; livestock feeding operations; livestock raised; volume of dairy, poultry, and livestock products
(5) Crop planting and livestock-breeding intentions
(6) Farm disposition of crops and livestock
(7) Processors' operations
(8) Prices received and paid by farmers
(9) Farm wage rates
(10) Farm employment
(11) Interest and tax payments
(12) Farm machinery information
(13) Land values
(14) Farm population and movement to and from farms

Statistics Computed from Primary Series:

To make these primary data easier to interpret and use, various kinds of series have been developed which might be described as statistical measuring-sticks.

These include the widely used index numbers of prices paid and received by farmers, of farm real estate values, volume of production, and farm output. They also include parity prices and livestock-feed price ratios. These measuring sticks, computed on the same basis for a period of time, show in brief form the gist of the meaning in many different groups of statistical series.

For instance the index of prices paid by farmers shows in one figure the present level of prices compared with a base period for 186 different commodities bought by farmers, plus interest and taxes. The index of prices received by farmers shows in one figure the present level of prices received for 165 commodities produced and sold by farmers also compared with a base period. When the index of prices received is divided by the index of prices paid, the resulting single figure, known as the parity ratio, shows how much farmers' products are worth in terms of purchasing power.

The types of statistics which are computed from the different primary groups described above are summarized in the following list. For some of these groups listed below, the primary statistics are combined with data from fact-finding agencies outside the Bureau to compute the series required.

(1) Supply and disposition of agricultural products
(2) Cash receipts and gross farm income, farm production expenses, and net farm income
(3) Financial condition of agriculture

(4) Retail value, farm value, and marketing margins of food products produced by farmers
(5) Total and per capita consumption of food products produced by farmers
(6) Feed supplies per animal unit
(7) Consumption of feedstuffs by various classes of livestock

Statistics about Farm Business and Farm People:

Although the above-listed statistical series cover a wide range of agricultural information there is demand from many sources for more statistics about farm business and farm people and for greater accuracy in the existing statistical series in this field. Steps are already being taken, insofar as resources permit, to collect more factual information by interview sampling and otherwise to serve as a basis for establishing new statistical series and for strengthening the Bureau's present series for wider use. The list which follows shows the subjects on which information, now inadequate or completely lacking, is most needed.

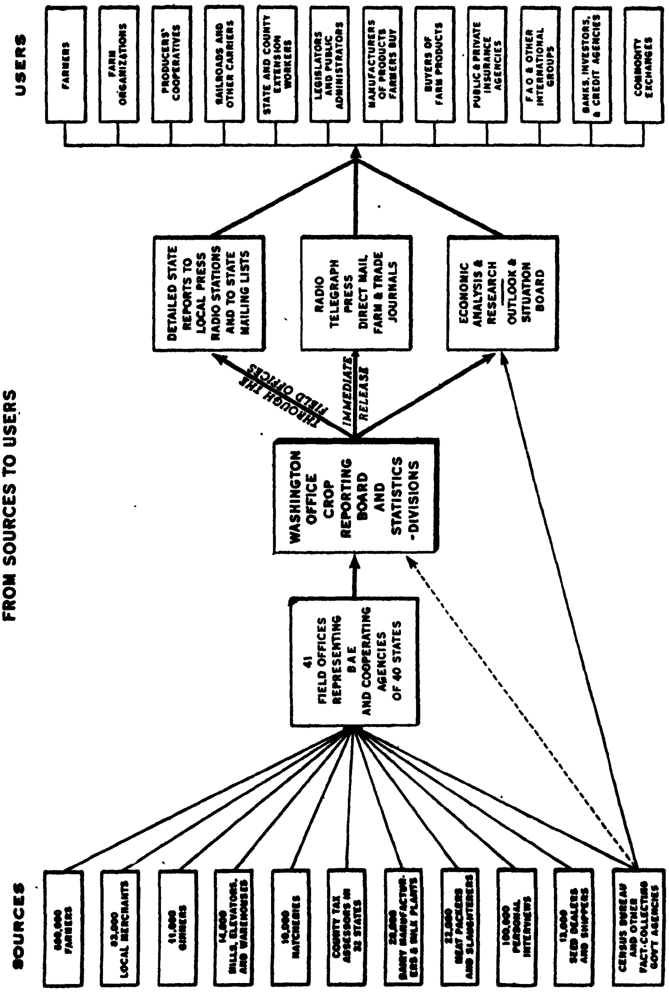
(1) Farm labor and farm wages
(2) Farm population and migration to and from farms
(3) Accidents on farms
(4) Sickness of farm people and medical and hospital facilities
(5) Farm home facilities
(6) Farm family living expenditures
(7) Farm production expenses
(8) Income from the sale of farm products
(9) Family income from sources other than farming
(10) Inventories of farm equipment, livestock, and crop stocks
(11) Net farm income
(12) Farm indebtedness, credit, and interest rates
(13) Land utilization
(14) Land ownership and tenure
(15) Land values and rents
(16) Values of farm buildings, new construction, and repairs
(17) Changes in farm management practices
(18) Methods and costs of marketing
(19) Farmers' marketing intentions
(20). Consumer preferences

The Bureau's statistical series showing changes in the farm business are not as complete as its statistics about farm products. As much of the basic information is obtained only once in the 5 or 10 years by the Bureau of the Census, many of these series are more valuable as historical data than for current use. Furthermore, the Bureau of Agricultural Economics has had to depend very heavily upon data from other

Federal, State, or private agencies. Much of this information, although adequate for special purposes for which it is collected, is difficult to use in the Bureau's official statistical series, especially those computed form primary series, which must be dependable and up-to-date, and must be maintained in a form which can be used for many purposes.

APPENDIX B

AGRICULTURAL ESTIMATES OF THE BUREAU OF AGRICULTURAL ECONOMICS FROM SOURCES TO USERS


USERS
FARMERS
FARM ORGANIZATIONS
PRODUCERS' COOPERATIVES
RAILROADS AND OTHER CARRIERS
STATE AND COUNTY EXTENSION WORKERS
LEGISLATORS AND PUBLIC ADMINISTRATORS
MANUFACTURERS OF PRODUCTS FARMERS BUY
BUYERS OF FARM PRODUCTS
PUBLIC & PRIVATE INSURANCE AGENCIES
F A O & OTHER INTERNATIONAL GROUPS
BANKS, INVESTORS, & CREDIT AGENCIES
COMMODITY EXCHANGES
DETAILED STATE REPORTS TO LOCAL PRESS RADIO STATIONS AND TO STATE MAILING LISTS
RADIO TELEGRAPH PRESS DIRECT MAIL FARM & TRADE JOURNALS
ECONOMIC ANALYSIS & RESEARCH
OUTLOOK & SITUATION BOARD
THROUGH THE FIELD OFFICES
IMMEDIATE RELEASE
WASHINGTON OFFICE CROP REPORTING BOARD AND STATISTICS DIVISIONS
41 FIELD OFFICES REPRESENTING BAE AND COOPERATING AGENCIES OF 40 STATES
SOURCES
600,000 FARMERS
93,000 LOCAL MERCHANTS
11,000 GINNERS
14,000 MILLS, ELEVATORS, AND WAREHOUSES
10,000 HATCHERIES
COUNTY TAX ASSESSORS IN 32 STATES
20,000 DAIRY MANUFACTURERS & MILK PLANTS
22,000 MEAT PACKERS AND SLAUGHTERERS
100,000 PERSONAL INTERVIEWS
13,000 SEED DEALERS AND SHIPPERS
CENSUS BUREAU AND OTHER FACT-COLLECTING GOV'T AGENCIES

This chart shows how information flows through the Bureau's statistical organization from the sources to the people who use it. The government's official agricultural estimates are based on reports from a cross-section of farmers and people who do business with farmers. They are checked against the complete enumerations made every five years by the Bureau of the Census, and against the reports of buyers, processors and carriers of farm products, whose operations are a measure of farm production. Besides their immediate usefulness to farmers and the general public, these estimates provide basic data required for analysis of farm income, prices, costs, parity price determinations, and prospective demand and prices for farm products.

The Bureau's limited staff of professional economists working in the field on cooperative projects with Land-Grant Colleges has collected a great deal of farm business data, but the information has been collected from farms representing a specific area and for the purpose of analyzing the special farming problems of that area. It is therefore difficult to combine this information into estimates for larger areas. The Bureau's research staff is now undertaking to develop farm management data that will show annual changes in systems of farming in income, expenses, and net income by types and sizes of forms for the more important farming areas. For this kind of statistical series, sample surveys are needed to provide data from representative farms on a comparable basis by areas, types of farms, and periods of time.

Résumé

Les premières statistiques agricoles publiées par le Département de l'Agriculture des Etats-Unis vers 160, fournissent des données sur l'état des récoltes et des estimation sur leur rendement. Ces statistiques ont été publiées pour répondre à la demande d'agricultures d'obtenir des informations propres à redresser les renseignements souvent faux ou erronés communiqués par les négociants de produits agricoles. Le domaine des statistiques agricoles s'est étendu au point qu'à peu près toutes les phases de la culture et de l'élevage, ainsi que les prix reçus et payeés per les agriculteurs et le marché du travail agricole sont maintenant compris dans les séries statistiques periodiques ou occasionnelles publiées par le Bureau d'Economie agricole. Les cultivateurs, les organisations de cultivateurs et les établissements commerciaux en rapports avec l'agriculture, ainsi que le Gouvernement Fédéral, sont devenus les plus importants usagers des statistiques agricoles.

Outre six divisions de statistique à Washington, connues sous la dénomination collective de "Estimation agricoles", le Bureau d'Economie Agricole compte 41 bureaux régionaux desservant chacun un ou plusieurs Etats. Trente-cinq de ces bureaux exercent leurs activités en collaboration avec des Départements d'Etat pour l'Agriculture ou des collèges d'agriculture. La coopération entre les Etats et le Gouvernement Fédéral évite les doubles emplois et permet la préparation d'estimations plus détaillées du point de vue géographique qu'il ne serait possible de le faire en employant seulement des fonds fédéraux.

Les plus importantes méthodes de réunion de renseignements statistiques sont: 1) énumérations complètes, dont les recensements agricoles fédéraux et dans 12 Etats, la préparation du coton, et l'industrie laitière

sont des exemples; (2) l'emploi de questionnaires envoyés et retournés par la poste; 3) observations sur place et contacts personnels par des fonctionnaires techniques; 4) mesurages et dénombrement effectués sur place par des fonctionnaires techniques, et 5) interviews d'échantillons de cultivateurs ou de préparateurs de produits agricoles, choisis d'avance suivant une méthode objective.

La plupart des estimations de récoltes et de cheptel faits par le Bureau sont basées sur des avis de changements reçus par la poste ou obtenus sur place par le personnel technique, changements qui sont mis en rapport avec les résultats du recensement quinquennal. Vu que beaucoup de renseignements reçus par la poste contiennent des erreurs systématiques significatives, on fait usage *autant que possible*, de méthodes de régression pour corriger de tels écarts s'ils sont constant. Toutefois, pour les prix reçus et payés parles cultivateurs, les renseignements reçus par la poste sont utilisés tels quels non seulement par le Bureau d'Economie Agricole mais aussi par le Bureau du Recensement pour l'estimation de là valeur de la production agricole.

Les besoins d'amélioration des statistiques agricoles les plus importants peuvent se résumer comme suit:

1) données plus compréhensives et plus fréquentes sur les changements technologiques (mécanisation, etc.) dans les opérations agricoles et dans les fermes;

2) statistiques plus fréquentes concernant le cheptel et l'élevage, y compris le bétail engraissé pour les marchés;

3) statistiques distinctes concernant la production "commerciale";

4) spécification régionale plus détaillée;

5) plus grande précision dans l'estimation de certains types ce phénomènes agricoles, et 6) presentation de données en termes de fréquences et autres caractéristiques d'une population.

Il sera probablement possible de satisfaire à ces desiderata de la manière la plus rapid et la plus économique en donnant de l'extension au programme actuel d'enquête par interviews d'échantillons choisis d'avance suivant une méthode objective, et une intégration de telles enquêtes avec des enquêtes par correspondance se rapportant aux mêmes personnes—enquêtes par correspondance "semi-contrôlées"—ou avec un rappel par interview des personnes qui n'ont pas répondu à l'enquête par correspondance—enquête par correspondance "contrôlée". La fréquence et la specification géographique, ainsi que les matières à inclure, dépendront des crédits affectés à ce travail. Ce but ne pourra être atteint à si peu de frais d'aucune autre manière.

THE SOCIAL SECURITY STATISTICS PROGRAM

by I. S. Falk
Social Security Administration
(United States)

Perhaps the most important characteristic of the social security statistics program—as compared with the other statistical programs which are included on the agenda for this session—is the fact that it takes its origin in social security program operations. Nearly all of the statistics which I shall review in this paper are designed to meet the specific needs of social security program operations, evaluation, and planning; with few exceptions, the basic data are derived directly from program operations.

The general interest and value of social security statistics, beyond their immediate significance for the social security programs, is of two kinds. In the first place, social security measures are sufficiently important in our economy to make statistics based on their operation of more use for general economic and social analyses than is usually recognized. In the second place, the size and continuity of the data provide an opportunity for the development of benchmark series which are useful in themselves and on which current estimates of more general use can be based.

In this brief paper I cannot attempt to discuss all the possible or significant uses of social security statistics. I shall describe only the major bodies of statistical data emanating from the social security programs. In general, I have selected for discussion those statistics which have significance beyond their direct or immediate administrative uses. I shall refer only incidentally to the development of derivative estimates for purposes of analysis, or to the general research program of the Social Security Administration. While our administrative needs largely determine what statistics are collected, it would require more space and time than is available to me to describe that program comprehensively.

The body of statistics with which I am here mainly concerned derives primarily from the operation of two social insurance programs (the national system of old-age and survivors insurance and the Federal-State unemployment insurance program), the three Federal-State programs of public assistance, and the Federal-State maternal and child health and welfare service programs. I shall also mention briefly the statistics derived from several related programs which are not administered by the Social Security Administration but which should not be omitted from any general review of social security statistics.

OLD-AGE AND SURVIVORS INSURANCE AND UNEMPLOYMENT INSURANCE

About three-fifths of all jobs in the United States are covered by the national system of old-age and survivors insurance and about one-half by the Federal-State unemployment insurances. The major exclusions are self-employment (including farmers and others in business for themselves), agricultural labor, domestic service, employment for nonprofit organizations, railroad employment, and government employment. Railroad workers and many government workers (Federal, State, and local) are covered by special systems which will be discussed briefly later. Substantially all workers in industry and commerce are covered by old-age and survivors insurance, and all such workers—except, in about three-fourths of the States, those employed by small firms—are covered by unemployment insurance.

Both insurance programs provide for the regular collection of contribution, which are used for the payment of benefits in specified circumstances. Under old-age and survivors insurance, the contributions are made jointly by employers and employees, and the benefits are payable to insured workers aged 65 and over who retire from covered employment, to their dependent wives and children under 18, and to specified surviving dependents of insured workers who die. Under unemployment insurance, the entire contribution is paid by the employers, except in 2 States which have employee contributions also; and the benefits are payable to individuals who recently earned wages in covered employment, who are able and available for work, but who are unemployed because of lack of work. Under both programs all persons in the specified employments are covered regardless of earnings; but contributions are required only on the first $3,000, earned in a year.

Two major categories of statistics result from the operations of each of these insurance programs: (1) statistics related to employer reporting of contributions, wages, and employment; and (2) statistics related to claims and benefit payments.

Wage and Employment Statistics

Under the old-age and survivors insurance program, each employer subject to the program files a quarterly tax report with the Collector of Internal Revenue (Treasury Department) for the district in which the business is located or, in case of multi-State employers, for the district in which the headquarters office is located. These reports are then sent to the Bureau of Old-Age and Survivors Insurance of the Social Security Administration for processing. In addition to information with regard to the amount of the contribution, each report contains a listing of all workers receiving taxable wages at any time during the quarter, the amount of taxable wages paid each worker, and the number of workers on the pay roll ending nearest the 15th of the last month of the quarter. Most employers having more than one place of business subdivide the information on taxable wages paid individual workers according to establishment or reporting unit. In general, the reporting unit is a single place of business engaged in a single industrial activity, although in some cases it is necessary to accept groupings of units so that reporting will not be too great a burden on an employer.

Under the unemployment insurance program, employers submit to the appropriate State unemployment insurance agency quarterly contribution reports which give the amount of total and of taxable wages paid during the quarter and the number of workers on the pay roll ending nearest the 15th for each month of the quarter. For multi-unit employers, the data are given for separate establishments or reporting units. Employment and wage data are compiled from these reports by each State agency, and are sent to the Bureau of Employment Security of the Social Security Administration in accordance with standard reporting procedures. Each State agency also obtains (for use in benefit determinations), either with the quarterly contribution report or at the time a worker is separated from employment, information on the taxable wages of individual workers.

Over the past ten years, repeated efforts have been made to simplify reporting requirements for employers by making it possible for an employer to submit the same contribution and wage report to both the Federal and State agencies. Differences in coverage and in other provisions of the Federal and State laws have prevented any general achievement of this goal, although many States have arranged their reporting forms so that the quarterly listing of names and wages of individual employees can be prepared, by use of carbon paper, simultaneously with the preparation of the corresponding part of the Federal report.

In order to assure maximum comparability of the wage and employment data obtained from these two sets of contribution reports, the Social Security Administration has, since about 1940, undertaken to promote the uniform industrial classification of employers (and establishments) reporting under the two insurance programs. The responsibility for the classification of new firms and for periodic resurveys of the nature of business of reporting firms is carried by the Bureau of Old-Age and Survivors Insurance, which sends the information to the State unemployment insurance agencies. If a State agency questions the classification assigned to a particular establishment, or has more recent information on the nature of its business activity, it informs the Bureau of Employment Security and the Bureau of Old-Age and Survivors Insurance and agreement is reached on the classification to be used in both sets of records.

Several different types and series of employment and wage statistics are derived from these records.

Monthly employment data. From the unemployment insurance reports, data are available on employment by month and on total and taxable wages by quarter, distributed by State and by industry within each State. The data are sent to Washington by the State unemployment insurance agencies on two different schedules. One is on an annual census basis, and includes data for all employers subject to the State laws during the year. The State-industry tabulations prepared by the Bureau of Employment Security from these State reports are released during the latter part of the succeeding year. In order to provide data on a more nearly current basis, the State agencies also prepare quarterly reports, due in Washington 4 months after the end of the quarter to which they relate, based on data from employers who reported for the previous and for the current quarters in time to meet the deadline date. From these "matched-employer" reports, estimates are calculated by relating the trend in the sample data to a base, or benchmark, derived from the last available annual report.

In addition to their use by the State unemployment insurance agencies and the Social Security Administration for administrative and program purposes, these monthly employment statistics are widely used throughout the government and by private groups interested in basic data on employment and wages. In spite of the coverage limitations of the data, they are the most comprehensive body of monthly employment and quarterly wage statistics available. Two major uses of the data by other agencies might be mentioned. The Department of Commerce use the wage figures extensively both in its computation of the national income and in its series of State Income Payments; and it uses the employment data in estimating full-time equivalent employment and average annual wages and salaries. The Bureau of Labor Statistics also uses the employment statistics periodically as a benchmark to which their current-sample national employment series is adjusted.

In order to correct for the omission of employment and wages in small firms not covered by unemployment insurance, estimates of such employment and wages are made on the basis of old-age and survivors insurance tabulations described below, and these are added to the unemployment insurance figures. There are a number of technical problems involved in the development of satisfactory benchmark employment data for the complete social security coverage. We have under way a number of special studies which we anticipate will result in improved benchmark employment and wage statistics.

All of the State unemployment insurance agencies tabulate and publish monthly employment data (for covered employment) by industry, for the State, and in some States for local areas within the State. Some of these agencies have also been interested in the preparation of current employment estimates for all nonagricultural employment. During the past year the Bureau of Employment Security and the Bureau of Labor Statistics have entered into a cooperative working arrangement for the development of employment and related data for the nation, the States, and local areas. In this project, the details of which are still being worked out, the major responsibility for the collection and publication of State and local area data will be placed, wherever possible, in the State unemployment insurance agencies. The same sample data will be used by the Bureau of Labor Statistics in preparing national series.

Business population statistics. From the quarterly employer tax returns, the Bureau of Old-Age and Survivors Insurance has available each quarter information on about 2 million employers. Counts of certain items on the reports are obtained each quarter in the course of the routine accounting operations. Once a year—in earlier years for the third calendar quarter and now for the first calendar quarter of the year—the Bureau tabulates data on the number of employers, their aggregate employment for the month of March, and their taxable pay rolls for the quarter, classified by various combinations of size-of-firm, industry, State, and county. Tabulations have also been prepared on an establishment or reporting unit basis, but these are subject to a degree of estimation (as is the aggregate employment item in the size-of-firm tabulations) since a substantial but decreasing number of employers fail to report the number employed as on the 15th of the month and none is required to report this figure for separate (multiple) establishments.

These tabulations are the only available source of statistics in intercensal years on size-of-firm for virtually all industry and commerce. The data are used by the Social Security Administration in analyzing the characteristics of employees covered under old-age and survivors insurance, the number of employers and amount of employment and wages included under old-age and survivors insurance but excluded from unemployment insurance. The statistics also have considerable general interest to anyone who is studying the characteristics of the business population, the relative importance of small and large firms in different industries, the concentration of employment and wages in firms of different size, and similar problems.

The Bureau of Old-Age and Survivors Insurance also tabulates each quarter, for the Office of Business Economics of the Department of Commerce, data on the number and characteristics of newly established and discontinued business by size-of-firm, State, industry, and circumstance surrounding the "birth" or "death." The business birth data are derived from the employer's application for an identification under old-age and survivors insurance. The quarterly business death statistics are procured from notices of discontinuance of business transmitted by the Collectors of Internal Revenue, supplemented by files of inactive firms in the Bureau of Old-Age and Survivors Insurance. The data are used primarily by the Office of Business Economics, in connection with the preparation of current estimates of business population and business turnover analyses.

Employers covered by unemployment insurance. Data are also available from the State unemployment insurance agencies on approximately 800,000 employers covered by the unemployment insurance program. All but one of the States have experience-rating provisions whereby contribution rates of individual employers may be varied from the standard rate. Annual tabulations classify employers under the State program by industry, tax rate, and size of annual pay roll. The data are used primarily for analyses of factors affecting employer's employment experience, variation in tax rates for firms of comparable type and size under different experience-rating systems, and similar problems.

Annual employee wage statistics. For 1946, annual wage credits amounting to about $67.8 billion, or about two-thirds of all civilian wage and salary payments in that year were recorded in the old-age and survivors insurance accounts of 49.5 million workers. In spite of the limitations that result from exclusion of certain employments and from lack of information on earnings in excess of $3,000 in a year, the records of individual earnings, year by year, under the old-age and survivors insurance program provide a unique body of statistical data of potentially wide value and usefulness.

Two sets of employee tabulations are prepared annually by the Bureau of Old-Age and Survivors Insurance. The first is based on a 3 percent sample of the individual employee wage records during the last calendar year for which wage recordings (in the old-age and survivors insurance accounts) have been completed. The tabulations vary somewhat from year to year, but in recent years have included distribution of workers by age, sex and color, annual taxable wages, number of calendar quarters in which the worker was employed during the year, State, and industry in which employed, and whether or not the worker changed his employer, State, or industry during the year. Annual tabulations of this kind are now available for each year from 1937 through 1945;

initial operations to prepare the tabulations for the year 1946 are now in process. Special tabulations have been made for some years for workers with wage credits in all four quarters of the year, who may be presumed to have worked primarily if not exclusively in covered employment and whose taxable earnings therefore closely approximate their total annual earnings (up to the maximum of $3,000 recorded for tax purposes). Special tabulations have also been made in a number of years for workers first employed in covered industry (since the beginning of the program in 1937) in the specified year. The majority of such workers are now persons newly entering the labor force, and their quarters in covered employment and taxable earning during the year are consequently low.

The Social Security Administration uses this time series of annual data in analyses of year-to-year changes in the employment and earnings characteristics of covered workers, in estimates of future administrative work loads and of levels of earnings under old-age and survivors insurance in analyses of alternative eligibility and benefit specifications for old-age and survivors insurance and unemployment insurance, and in studies of disability and health insurance.

Work history data. Annual data represent only single year's employment and earning experience of workers covered by the insurance programs. Eligibility for benefits—and more particularly the amount of benefits under old-age and survivors insurance—are dependent on the work and wage patterns over the working lifetime. The Bureau of Old-Age and Survivors Insurance has consequently developed a second series of tabulations—based on a 1 percent sample—showing the work and wage patterns of workers in covered employment at any time since the inception of the old-age and survivors insurance program in 1937. These individual work history statistics provide information on the number of persons who have contributed to the insurance system and the cumulative amount of their contributions, as well as the number who meet the requirements of insured status; they also provide distributions of workers by age, sex, and color, by number of years and specific years in covered employment, cumulative number of quarters of employment, annual and cumulative wages.

Economists and statisticians will immediately recognize the great potential value and usefulness of both the annual earnings and the individual work-history tabulations. The two series are now available for a period of nine years. We have begun to analyze the trend and patterns of employment and earnings revealed by the series, with special relation to their implications for the social security program. The opportunities for more general analysis of the data already published have only begun to be exploited.

It should be mentioned that many State-unemployment insurance agencies now prepare for their own use tabulations, from all or a sample of their wage records, of wages earned by individual workers during the calendar year, cross-classified by the amount of such wages which were earned in each individual's most remunerative calendar quarter. These data are, for most States, basic factors in determining benefit rights. Nearly one-half the States are now sending copies of these tabulations to Washington on a voluntary basis. All State agencies estimate each year, as part of the regular statistical reporting program, the number of different workers earning wages under the program during a given year.

Benefit Statistics

From the point of view of social security program administration, the employment and wage statistics provide information on the population at risk—that is, the potential beneficiaries of the programs. Statistics of benefit operations give us information as to the persons for whom the insured risk has materialized and the way in which the program affects them. The benefit statistics, like the employment and wage statistics, reflect the operations of the economy and can be used to trace some of the interactions between the social security program and other factors affecting the Nation's economy.

Old-age and survivors insurance benefit statistics. Two main types of statistics have been developed relating to old-age and survivors insurance beneficiaries. One is obtained directly from claim and benefit records and includes tabulations of: (1) insured workers, on whose wages either retirement or survivor claims were awarded; (2) individual beneficiaries and their benefits; and (3) beneficiary families and their benefits. Tabulated data include information with respect to marital status, age, sex, race, insured status, type and amount of benefit, data of entitlement to benefits, and date and cause of termination of benefits.

The extent to which old-age and survivors insurance benefits meet the needs of beneficiaries and fulfill the objectives of the Social Security Act depend to a very large degree upon the additional resources which beneficiaries have. The Bureau of Old-age and Survivors Insurance has consequently made a number of field studies of retired and survivor beneficiaries in selected localities. In these surveys information has been obtained on the income, employment status, living arrangements, and assets and liabilities of beneficiary groups and, in more recent surveys, on the cost of medical care for beneficiaries during the survey year, number of children outside their households, and amount of payments covering joint expenses in households where beneficiaries and relatives lived together.

While these statistics are primarily useful in connection with the administration and development of old-age and survivors insurance, they provide a significant body of information about members of two groups in our population—the aged and the widows and orphans—which is of considerable potential interest to economists and sociologists and to legislators and others concerned with the welfare of these groups.

Statistics related to unemployment insurance benefit operations. In most countries which have unemployment insurance systems, statistics based on unemployment claims data have long been used as indicators of changes in economic activity.

Such statistics have been highly developed in this country. Data on the volume of various types of claims filed are reported to Washington by the State unemployment insurance agencies on a weekly and monthly basis providing an indication of the location, extent, and duration of unemployment in employments covered by State unemployment insurance laws. Current claims data from the Railroad Retirement Board with respect to the railroad unemployment insurance system, and from the Veterans Administration with respect to veterans, readjustment allowances, are combined with the State unemployment insurance data and issued weekly by the Bureau of Employment Security to provide the most current series available relating to unemployment in the nation and each State. These data are used widely by government and private organizations interested in the current trends of unemployment.

Although the unemployment insurance claims figures do not reflect all unemployment (they exclude workers unemployed beyond the benefit period, as well as those in industries not covered by the insurance program), they provide a current and sensitive index of industrial and commercial unemployment. Moreover, they provide a current and sensitive index of industrial and commercial unemployment. Moreover, they provide geographical distributions of unemployment—by State and by local areas. Current unemployment indicators for small geographic areas are invaluable in identifying problem areas in their early stages. Trends in new claims reveal the pattern of the impacts of new unemployment among covered workers and, when compared with data on waiting period and compensable claims, they provide an indication of the extent to which claimants are immediately reemployed or are employed before reaching compensable status. Claims data by local office areas are useful in analyzing local labor market conditions.

More detailed information as to claimants and beneficiaries is available on a monthly, quarterly, and annual basis. Statistics are reported monthly by the State unemployment insurance agencies on the number of weeks of unemployment for which payments were made, the amount of such payments, the number of first payments issued to claimants within their benefit years, and the number of instances wherein claimants exhaust all benefit rights to which they were entitled. From these data are derived figures on the average benefit amount, the number of different individuals receiving benefits during a year and the proportion of unemployed workers who exhaust benefits before they are reemployed.

Quarterly reports indicating the potential duration of benefits payable to new claimants who are eligible for benefits, by weekly benefit amount, show the extent of protection against wage loss offered to current claimants under existing provisions of State laws, and reflect on a reasonably current basis the effect of changes in benefit provisions of State laws and in earnings levels. After claimants have completed their benefit period, reports are submitted comparing, for claimants with specified weekly benefits, the actual duration of benefits with the potential duration.

There are five States which now have provisions in their laws for paying benefits that vary in amount according to the number of the worker's dependents and these States are submitting statistics showing the effects of this provision. The data indicate the number and type of dependents on whose accounts increased benefits are payable, the amount of the increase, and something as to the relative earnings level of workers with and those without dependents (as defined in the legislation).

All State unemployment insurance agencies submit data quarterly on the number of new claims allowed and disallowed, the length of time elapsed between the end of the benefit period and the date of payment, the number of disqualifications (classified by disqualification issue and severity), and the number of appeals (classified by issue, source, effect on claimant, and time required to dispose of them). These statistics serve principally to reflect the characteristics of program operations.

I shall not take the time to list a number of other statistical series that are used primarily for administrative purposes. I might, however, mention that two States, Rhode Island and California, now have temporary disability (cash sickness) insurance programs which are administered by the State unemployment compensation agency in close conjunction with the unemployment insurance programs. Data on claims, benefits, and beneficiaries under these temporary disability programs are reported voluntarily by the State and are published with other social security data in the *Social Security Bulletin*.

PUBLIC ASSISTANCE STATISTICS

The public assistance programs are State-administered programs for making money payments to individuals who have been found to be in need—that is, to have insufficient resources on which to live according to standards established by the States. The Federal Government matches, up to specified amounts, State expenditures for assistance to needy aged persons, to the needy blind, and to dependent children. There are no Federal grants-in-aid for general assistance (*i.e.*, aid to persons who are not within these specified categories). Payments to needy persons outside of the three special categories are financed wholly by the States and localities.

States having approved plans for the three special types of public assistance report periodically to the Bureau of Public Assistance in the Social Security Administration on the operations of those programs. All of the States and 19 large cities also make supplementary (voluntary) reports relating to general assistance.

Monthly data are available for old-age assistance, aid to the blind, aid to dependent children, and general assistance, and for all 51 jurisdictions (the 48 States, the District of Columbia, Alaska, and Hawaii) on the number of persons receiving assistance, the volume of assistance payments, and (by derivation) the average monthly payment per recipient. Data are also available on a monthly basis, on the number of applications received and acted upon during the month, the number of applications pending at the beginning and end of the month, the number of cases carried forward at the beginning and end of the month. These data provide a rough current index of the impact of changing economic conditions on the circumstances of families at the margin of subsistence. The index is less sensitive than it might be because of the large volume of unmet need which we know, on the basis of special studies and analyses, exists even in relatively prosperous times. Nevertheless, when combined with quarterly data on reasons for opening cases and semiannual data on reasons for closing cases the figures reveal fairly clearly the effects of changing employment opportunities for recipients or members of their families, of changing costs of living and of the operation of the insurance programs. County data on number of recipients and total assistance payments which are reported semiannually, and data relating to general assistance in 19 large cities, make possible analysis of local area differences and trends.

I shall not discuss all the other statistical series relating to public assistance which have primarily administrative uses. I would mention only one series—the semiannual data for all States on expenditures for public assistance payments and administration, showing amounts from

Federal, State, and local funds. In addition, I might refer to several special statistical studies now under way or recently completed by the Bureau of Public Assistance and State public assistance agencies which volunteered to participate in the studies. A study of recipients of old-age assistance in 21 States, which has just been completed, provides information on the age, sex, and other characteristics of the recipients, their living arrangements, the amount which the public assistance agency had established as the amount the recipient needed to live on, the amount of the assistance payment, and the amount and sources of income other than assistance. Also completed recently is a special study of the causes of blindness, conducted by 20 States. A study of the medical aspects of public assistance which is now under way will provide statistics on the volume and costs of medical services of various types supplied to needy persons by selected local agencies over a 6-month period.

STATISTICS RELATING TO MATERNAL AND CHILD HEALTH AND WELFARE SERVICES

Special health and welfare services for mothers and children are available in many localities throughout the country under the Federal State programs administered by the Children's Bureau in the Social Security Administration. Federal funds help finance services provided by State health agencies through prenatal clinics, well-baby and child health clinics, or public health nursing services for infants, preschool children, and school children, and diagnostic and treatment services for crippled children. They also help finance special services provided by State and local welfare agencies to children—especially dependent and neglected children and those in danger of becoming delinquent—who are in their own homes, in institutions, or are in need of day care, or adoption. The Children's Bureau also administers funds (at present on a postwar liquidation basis) which are used by State health agencies to pay for maternity and infant care for families of enlisted men in specified grades of the armed forces.

Periodic reports by the State health and welfare agencies provide information on the volume and kinds of services which are being made available to these special maternal and child groups in the population. Data are reported quarterly on the volume and types of maternal and child health services, services to crippled children, and the cost of medical care for wives and infants of enlisted men provided under programs administered or supervised by State health agencies, and the number and living arrangements of children served by State and local welfare agencies. Annual reports include more detail on the characteristics of children receiving services, and on the amounts spent and the personnel engaged in providing such services.

Statistics reported under these programs, when used alone or in relation to other statistical data, give some idea of the accomplishments and also of unmet needs and gaps in the services currently provided. For example, the annual reports on personnel in the child welfare program indicate that child welfare workers are concentrated in a small number of States and in urban areas out of proportion to the child population in these States and areas, while other sections of the country with large proportions of the child population have few specialized child welfare workers. Similarly, in some States the need for the expansion of the child welfare program to include service to children in their own homes is indicated by unduly high proportions of children receiving service in foster-family homes or in institutions.

State agencies administering the crippled children's program maintain registers of crippled children as well as records of services provided. In 1946 approximately 108,000 crippled children received diagnostic and treatment services at clinics financed in whole or part by Federal funds; while at the end of that year 442,000 were on the crippled children's registers maintained by the State agencies administering these funds. This latter figure includes, of course, only crippled children known to the agencies, not all those suffering from such disabilities. Work is now in progress on a standard classification of crippling conditions to be used by State agencies in reporting on services provided to crippled children under these programs.

Although the program is now being liquidated, reference may be made to the statistics relating to the operation of the emergency maternity and infant care program. Between April 1943 and June 1947, maternity care had been authorized for over 1,203,000 wives, and medical, hospital, and related services during the first year of life for approximately 217,000 infants, of men in the lowest four pay grades of the armed services. Monthly statistics are available on cases authorized and completed and quarterly statistics on types of services and costs of care provided under the program.

STATISTICS FROM RELATED PROGRAMS

This review would be incomplete if it did not refer, however briefly, to a few of the more important statistical series derived from social security and related programs not administered by the Social Security Administration.

Railroad and related transportation workers in this country are now covered by a social insurance system providing retirement, survivor, disability, sickness, and unemployment benefits. From the contribution reports of employers covered under the railroad social insurance program, data are available on total and taxable wages (wages up to $300 a month) by quarter, and on the number of covered workers by month. The Railroad Retirement Board prepares annual tabulations of the number of workers covered under the railroad program, their annual taxable earnings and months of service in railroad employment, by age and by occupation. The Social Security Administration and the Railroad Retirement Board have in the past made a number of special statistical studies of workers having earnings under both programs during the course of a year.

Statistics are also available monthly on the number of beneficiaries and amount of benefits paid under the railroad program. Reference has already been made to the use of weekly data on unemployment benefit claimants under the railroad program in the insured unemployment series prepared by the Bureau of Employment Security in the Social Security Administration.

Most Federal Government workers are now covered under the Civil Service Retirement System. The Civil Service Commission publishes annual statistics on the number and characteristics of persons receiving retirement and disability benefits, or joint-survivor annuities and lump-sum death payments.

The veterans' program may be regarded as a social security program for a special group in our population. I shall not attempt to review the available statistics relating to veterans' benefits. It may be of interest to you to know that the Bureau of Research and Statistics in the Social Security Administration has developed a series of comparable data on the number of beneficiaries and amount of benefits, by month, under the railroad, civil service, and veterans' benefit programs as well as under old-age and survivors insurance, State unemployment insurance, and the two State temporary disability programs. To complete the series, we prepare annual estimates of benefits and beneficiaries under the more than 1,700 State and local employees' retirement systems and of payments under the Federal and State workmen's compensation programs. We have not, as yet, been able to develop a method of estimating the number of persons receiving payments under more than one program, and thus to arrive at an unduplicated count of the number of individuals involved. The composite series, in conjunction with data on recipients of public assistance, does give us an approximate measure of the extent to which various groups in the population are reached by the existing social security and related programs.

Summary

Because the social security program includes within its compass both persons actively at work and persons who are temporarily or permanently out of the labor force, statistics derived from the operations of that program can be used to observe and analyze many aspects of the social dynamics of our economy. The main inadequacy of the statistics for purposes of general economic analysis stems from one of the major limitations in the existing social insurance program—its failure to cover workers in all employments. For the groups covered, the social security statistics provide information of a kind not otherwise available, and of a degree of accuracy not easily achieved through other methods used in the collection of large-scale economic or social statistics. While all of the basic data rest upon a system of operating reports for the entire population covered by the programs, many of the tabulations are prepared on a sample basis.

As indicated earlier, the major emphasis in the development of the statistics is upon their use in the administration of the social security program, and in studies and analyses of the effectiveness of that program and the extent of unmet social security needs. In order to carry out the responsibility imposed by the Social Security Act for "studying and making recommendations as to the most effective methods of providing economic security ......," we must, of course, go beyond the statistics derived from the operations of existing programs.

In studies of the social security needs and the problems of extending coverage to groups now excluded, we have turned to data from the census of population, from special studies of farmers and agricultural workers, and from other sources. In some cases, we have ourselves undertaken field studies of such groups as domestic workers or the urban self-employed; but, so far as possible, for supplementary statistics we rely on data collected by other agencies.

In our studies of disability and health insurance, also, we rely on data from other sources—studies of the incidence and characteristics of disability under private benefit programs or under foreign social insurance systems, the results of special surveys, the operations of many kinds of public institutions, the data on medical personnel and facilities collected by the Public Health Service or private organizations, and similar data. In this area of study, we have also found it necessary to compile some information ourselves, such as periodic reports on the number of prepayment medical care organizations, their membership, and the types of service provided.

Our studies of the financing and the fiscal aspects of the social security programs rely partly on cost data derived from social security program statistics, partly on data from tax yields, national and State income estimates, and similar information from other sources. It would take me too far afield to discuss the uses we make of such data in analyses of relative fiscal capacity of States and localities or of the impact of social security contributions and benefit payments on the economy as a whole, on States, on economic groups, etc.

Although our major interest is in the use of social security statistics for administrative and other program purposes, the Social Security Administration has always been alert to the other important uses of social security data, and has attempted to take into account other public interests than its own. Limitations of budget and the fact that we are not, under existing legislation, able to prepare tabulations or do other special work on a cost reimbursement basis except for Federal agencies, make it impossible for us to respond to many requests for special tabulations or listings of information that could easily be obtained from our basic statistical records.

In closing, I would again call the attention of statisticians and economists to the growing body of data available as a result of our regular periodic tabulations. Most of these data are published regularly in the *Social Security Bulletin* or *Yearbook*; others appear in special reports. Even without the elaborations that would be desirable for many purposes, this body of data affords many opportunities for analyses of a wide variety of economic and social problems.

Résumé

Les statistiques de sécurité sociale découlent principalement de la mise en pratique de deux programmes d'assurance sociale (le système national d'assurance-vieillesse et assurance-veuves et orphelins, et le programme Fédéral-d'Etat (Federal-State) d'assurance-chômage, des trois programmes Fédéral-d'Etat d'assistance publique, et le programme Fédéral-d'Etat d'aide maternelle et infantile.

Deux catégories principales de statistiques résultent de l'application de chacun des programmes d'assurance: les statistiques, basées sur les données fournies par les patrons, concernant les primes, les salaires et le personnel et les statistiques ayant trait aux demandes d'indemnités et aux indemnités payées'.

Des renseignements relatifs aux primes d'assurance-chômage on obtient des chiffres sur le chômage par mois et sur les salaires totaux et imposables par trimestre; répartis par Etat et par catégorie d'industrie. Malgré le fait que les données présentent certaines lacunes, elles forment l'ensemble le plus compréhensif de statistiques mensuelles de l'emploi et de statistiques trimestrielles des salaires dont on dispose.

Des statistiques du nombre d'établissements industriels et commerciaux et des tableaux annuels d'employés sont préparés au moyen des informations sur la taxe à payer par les patrons suivant le programme d'assurance-vieillesse et assurances-veuves et orphelins. Les tableaux sur les établissements comprennent, pour un trimestre de chaque année, des chiffres sur le nombre de patrons, leur personnel et les salaires payés, classés suivant la grandeur de l'établissement, la catégorie d'industrie, l'Etat et le district et, pour chaque trimestre, des renseignements sur le nombre et les caractéristiques d'entreprises nouvelles ou ayant cessé d'exister. Les tableaux annuels relatifs aux personnes employées donnent e.a. des renseignements sur la classification des travailleurs par âge, sexe, race, trimestres pendant lesquels on a été occupé, l'Etat l'industries où on est employé, changements de salaire et d'occupation pendant l'année.

A l'aide des rapports relatifs aux demandes d'indemnité et aux indemnités payées d'assurance-viellesse et assurance-veuves et orphelins on prépare des tableaux sur les travailleurs assurés, les bénéficiaires individuels et les familles-bénéficiaires. Ces tableaux comprennent des renseignements sur l'état matrimonial, l'âge, le sexe, la race, l'état d'assurance, le type et le montant de l'indemnité, la date de l'entrée en jouissance et les date et cause de la cessation de l'indemnité. Ces chiffres, résultant de l'application du système d'assurance, sont complétés par des enquêtes spéciales qui fournissent des renseignements sur la situation économique de groupes de bénéficiaires dans certaines villes.

Des chiffres sur l'importance de différents types de demandes d'indemnité reçues sont communiqués à Washington par les agences d'assurance-chômage de l'Etat, hebdomadairement et mensuellement, ce qui permet d'obtenir des renseignements sur la répartition géographique, l'extension et la durée du chômage assuré. Des chiffres plus détaillés ur les demandes d'indemnité et les bénéficiaires sont disponibles par trimestre et par an.

Les Etats qui ont des plans approuvés d'assistance-vieillesse, d'assistance aux aveugles et à des enfants, soumettent mensuellement des rapports sur l'application de ces programmes, et tous les Etats et 19 grandes villes envoient volontairement des rapports sur l'assistance générale. Ces rapports contiennent des chiffres sur le nombre des personnes assistées, les paiements, les demandes d'assistance et sur le nombre des cas nouveaux et des cas ayant pris fin. L'Administration de Sécurité Sociale procède aussi à des études sur les caractéristiques économiques et sociales des bénéficiaires.

Le Bureau dejeunesse de l'Administration de Sécurité Sociale rassemble trimestriellement des chiffres sur les activités relatives aux programmes concernant l'aide maternelle et infantile, les soins aux enfants boiteux, et le nombre des enfants profitant des services d'assistance de l'Etat et locaux ainsi que les arrangements pris en ce qui concerne ces enfants. Des rapports annuels contiennent des renseignements plus détaillés; on y trouve des données sur les caractéristiques des enfants assistés, sur les dépenses et sur le personnel.

La plupart des chiffres sont publiés dans le "Social Security Bulletin" ou le "Yearbook"; d'autres sont donnés dans des rapports spéciaux.

Discussion

Mr. Stone:

Dr. Falk has given us a most impressive picture of the wide range of statistical data which the Social Security Administration is able to give us now. He has called attention to two limitations: one is that they give us the figures only on the industries which are covered by the law, which cuts out a significant section of employees; the other is the limitation to the first $3,000 of earnings a year. At the time when the law was adopted, it is quite conceivable that it was a perfectly reasonable limitation because very few wage-earners earned more than $3,000. But with the considerable increase in earnings which has taken place in the last few years a large body of wage-earners are earning considerable more. I wonder whether Dr. Falk could exercise his ingenuity in getting the actual earnings figures in spite of the limitations so far as the payment for security is concerned.

Dr. Falk:

May I say that the comments are well taken. We have been trying in various of our bureaus to work out methods whereby, through estimates of various kinds, the limitations of the data may be reduced to a minimum.

As to the $3,000 limit, we assume that when the Congress gets around to a review of the social insurances, the limit will be raised to take into account the upward trend, or higher level, of wages. But there will then remain a new ceiling. Let me point out that in the unemployment insurance statistics, we do get reports of total payrolls for those who are covered by unemployment insurance. We estimate regularly aggregate earnings over $3,000 in industries covered by old age and survivors insurance. For special research purposes, approximate estimates are made from time to time of the presumed distribution of the earnings in excess of $3,000 by reference to other series.

The limitations that result from coverage present much more complex problems and constitute one of the most important reasons why persons using these series must take into account the substantive characteristics that are behind the data. The real remedy for that problem is, of course, a substantive one. We must wait on expansion in the coverage of the insurance systems by Congress; an amendment to our laws which we have frequently recommended and which we hope very much will come in the near future.

STATISTICAL METHODOLOGY SECTION

Tuesday, September 9, at 4:00 p.m.

CHAIRMAN :

R. A. Fisher

ANALYSE DES LIAISONS DE PROBABILITE

par Georges Darmois
Université de Paris (France)

On peut ainsi nommer tout le corps des recherches qui visent à la représentation, éventuellement à l'explication, de liaisons stochastiques par l'intermédiaire de ce qu'on pourrait appeler en général des variables aléatoires cachées, et que, dans la terminologie de l'analyse factorielle, on qualifie justement de facteurs.

Ces variables cachées sont, dans les recherches classiques, indépendantes en probabilité, et les variables observées sont supposées en être des fonctions, mais ce sont, disons les plus simples de la hiérarchie des fonctions. En fait, ce sont des combinaisons linéaires de facteurs indépendants; dans la notation habituelle, ce sont des combinaisons pondérées de variables centrées et normées (c'est-à-dire $E(a) = 0$; $E(a^2) = 1$, si a est un facteur).

A proprement parler, l'analyse factorielle classique ne vise qu'à reproduire au mieux, par des conditions d'optimum portant en particulier sur le nombre des facteurs, les moments du second ordre de la loi de distribution étudiée, ou encore, ce qui revient au même, à retrouver au mieux les coefficients de corrélation.

On peut, se plaçant d'abord à un point de vue théorique, étendre le problème de l'analyse factorielle en gardant la notion générale de l'indépendance des variables cachées, mais en permettant aux variables aléatories étudiées d'en être des fonctions plus générales que des fonctions linéaires.

On peut ainsi représenter, comme nous le verrons, des lois de probabilité qui ne rentreraient pas dans la classe, théoriquement assez restreinte, de celles que fournissent les schémas linéaires.

Les fonctions des facteurs dont nous allons nous occuper ne sauraient toutefois garder toute leur généralité. En effet, toutes les variables aléatoires d'un même champ, ou comme le dit M. Fréchet, correspondant à une même catégorie d'épreuves, sont déterminées quand l'épreuve est faite, et peuvent ainsi être considérées comme des fonctions d'une même variable, qu'on peut appeler l'aléatoire fondamentale. Or, ces fonctions d'une même variable peuvent présenter tous les degrés de la dépendance stochastique. Pour cette raison, nous donnerons à ces fonctions des facteurs certaines propriétés de régularité.

Par exemple, si une loi de probabilité à deux variables x_1 et x_2 peut être représentée avec deux facteurs α_1 et α_2 par les équations ;

$$(1) \qquad x_1 = f_1(\alpha_1, \alpha_2) \qquad\qquad x_2 = f_2(\alpha_1, \alpha_2)$$

nous ferons sur f_1 et f_2 les hypothèses nécessaires pour que α_1 et α_2 puissent être pris comme de nouvelles coordonnées du point aléatoire (x_1, x_2), et qu'on puisse passer par les méthodes classiques du calcul différentiel d'une densité de probabilité à l'autre. Et nous parlerons des courbes $\alpha_1 = C^{te}$, $\alpha_2 = C^{te}$, etc.

Nous nous bornerons ici aux distributions absolument continues. Soit Soit alors, pour les variables x_1 et x_2 la probabilité élémentaire:

$$a(x_1)dx_1 . \quad a^{x_1}(x_2)dx_2 .$$

$a(x_1)dx_1$ est la probabilité elémentaire marginale de x_1, $a^{x_1}(x_2)dx_2$ est la probabilité élémentaire liée de x_2.

Il y aura liaison stochastique si $a^{x_1}(x_2)$ dépend effectivement de x_1. Nous allons dans ce cas, chercher une solution du type (1), mais une solution particulière en ce sens que nous prendrons comme premier facteur x_1

$$x_1 = \alpha_1 \qquad\qquad x_2 = f_2(x_1, \alpha_2)$$

et résolvant la deuxième équation par α_2 nous aurons:

$$\alpha_2 = \varphi_2(x_1, x_2)$$

on voit immédiatement qu'on doit avoir:

$$(2) \qquad \frac{\partial \varphi_2}{\partial x_2} = a^{x_1}(x_2)$$

$$(3) \qquad \varphi_2 = F^{x_1}(x_2) + C(x_1)$$

$Fx_1(x_2)$ est la fonction de répartition de x_2 liée par x_1, $C(x_1)$ arbitraire. Nous avons d'abord la solution particulière $\alpha_2 = F^{x_1}(x_2)$

On voit qu'on peut introduire en espace à trois dimensions x_1, x_2, Z.

La surface représentée par l'équation $Z = F^{x_1}(x_2)$ a pour sections par les plans $x_1 = C^{te}$ les courbes de répartition de x_2 liée par x_1. Les sections par les plans $Z = C^{te}$ se projettent sur le plan des $x_1\, x_2$ suivant les courbes $\alpha_2 = C^{te}$ qui, associées aux droites $x_1 = C^{te}$, constituent un système de coordonnées curvilignes indépendantes en probabilité.

La construction de ces courbes $\alpha_2 = C^{te}$ peut s'obtenir par une méthode géométrique. Si l'on considère une bande infiniment étroite x_1, $x_1 + dx_1$ dans laquelle x_2 est une aléatoire liée, la probabilité totale, disons la masse totale de cette bande est $a(x_1)dx_1$. Et le rapport à cette mass totale de la masse partielle telle que l'ordonnée soit inférieure à x_2 est justement $F^{x_1}(x_2)$, c'est à dire α_2. Les courbes $\alpha_2 = C^{te}$ qu'on obtient par cette construction seront dites associées aux droites $x_1 = C^{te}$ ou mieux en raison de la symétrie, que ces deux familles constituent un réseau de courbes associées.

Remarque. Nous pouvons observer que l'indépendance des variables aléatoires x_1 et α_2 est évidente, car pour une valeur fixée quelconque de x_1, la loi de probabilité de α_2 est toujours la même. α_2 est répartie uniformément entre zéro et un.

Cette remarque rend intuitives les généralisations, en particulier celle-ci:

Si l'on remplace par une famille de courbes la famille $x_1 = C^{te}$, on voit que la construction précédente, faite cette fois sur la bande constituée par deux courbes infiniment voisines fournira des courbes associées. On voit donc qu'il existe une infinité de réseaux de ce genre. En particulier, à une famille quelconque de droites parallèles, on peut associer une famille bien déterminée.

On sait que pour une loi de Laplace-Gauss, cette famille associée serait encore composée de droites, et que les directions de deux familles sont conjuguées dans l'ellipse indicatrice. Disons simplement que cette propriété est caractéristique de la loi de Laplace-Gauss, en ce sens que si deux réseaux associés différents sont composés de droites, la loi de probabilité est nécessairement de Laplace-Gauss.

LOI DE PROBABILITÉ A TROIS VARIABLES

Considérons une loi élémentaire de la forme:

$$a(x_1, x_2)dx_1,\ dx_2.\ A^{x_1 x_2}(x_3)dx_3$$

Si l'on prend: $\alpha_3 = F^{x_1\, x_2}(x_3)$ α_3 a pour valeur la fonction de répartition de x_3, liée par le couple x_1, x_2. C'est une variable indépendante du couple, et les surfaces $x_3 = C^{te}$ seront dites associées à la congruence des droites $x_1 = C^{te}$, $x_2 = C^{te}$.

Si maintenant, considérant la loi à 2 variables x_1, x_2, on fait la construction de l'un des réseaux de fonctions associées, on aura déterminé trois facteurs

$$\alpha_1(x_1, x_2) \qquad \alpha_2(x_1, x_2) \qquad \alpha_3(x_1, x_2, x_3)$$

Et cette construction peut se faire à partir d'une congruence qui généralise celle des droites $x_1 = C^{te}$, $x_2 = C^{te}$. Il suffit en effet dans le tube infiniment petit qui entoure une courbe de la congruence, de déterminer en un point le rapport de la masse partielle s'étendant de l'infini au point considéré à la masse totale, on aura la variable α_3. Les variables α_1, α_2 sont celles de deux surfaces associées de la congruence.

On voit que nous avons déjà démontré quelques propositions générales, valables pour les lois absolument continues:

a) Il existe une infinité de décompositions de type (1) correspondant à des réseaux de courbes associées.

b) Si l'on applique la même méthode à des variables indépendantes on voit qu'il existe une infinité de couples $f_1(\alpha_1 \; \alpha_2)$, $f_2(\alpha_1 \; \alpha_2)$ qui, partant de variables indépendantes, redonnent des variables indépendantes. Cette proposition était bien connue pour les lois de Laplace-Gauss.

Autres décompositions

Considérons par exemple une famille de courbes fermées qui grandissent et s'enveloppent à partir d'un point O, et une courbe Γ_0 formée d'un axe parti de O et allant à l'infini que fixera sur chaque courbe fermée une origine Ω.

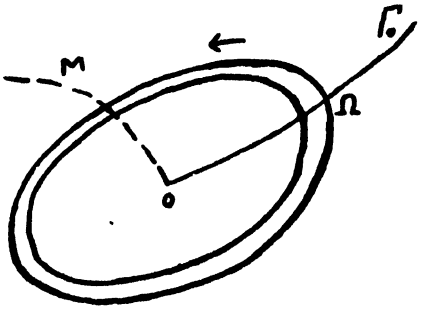


Fig. 1

Si l'on marque un sens sur les courbes fermées, on pourra, dans chaque anneau infiniment petit, définir un point M par le rapport α_2 à la masse totale de l'anneau de la masse partielle comprise entre Ω et M. Les courbes $\alpha_2 = C^{te}$ seront analogues à Γ_0: si α_1 est le paramètre des courbes fermées, α_1 et α_2 seront encore deux variables aléatoires indépendantes.

Il est remarquable qu'à une famille $\alpha_1 = C^{te}$, on puisse associer une infinité de familles, en choisissant arbitrairement Γ_0. Par exemple, il est bien connu que, dans une loi de Laplace-Gauss, les éllipses et les droites issues du centre constituent un réseau d'associés. Mais on voit qu'on peut remplacer les droites par une famille de courbes, également issues du centre. Cet ordre d'idées va nous amener à l'interprétation de la solution générale

$$\varphi = F^{x_1}(x_2) + C(x_1). \tag{3}$$

Mais pour voir les choses bien clairement, il est préférable de faire d'abord une transformation. Les associées au sens particulier des droites $x_1 = C^{te}$ étant prises comme coordonnées, on peut toujours supposer les variables x_1, x_2 indépendantes, et variables entre zéro et un. La loi primitive est alors représentée sur le carré de côté un. Appelons les variables x et y. Soit alors une courbe Γ_0. A tout point M, correspond une fonction que nous pouvons encore appeler α_2, et qui représente dans chaque bande le rapport à la masse totale de la masse partielle comptée à partir de Ω. Pour le point M, $\alpha_2 = \Omega M$. Pour le point N c'est $\Omega' H + KN = 1 + \Omega' N$. La courbe $\alpha_2 = C^{te}$ se compose de deux arcs déduits par translation des deux arcs AB, BC de Γ_0. Mais on voit que dans la représentation primitive les points P et Q ont pour images des points à l'infini (ou plus généralement des points de la frontière). Par exemple, si Γ_0 est la première bissectrice, α_2 a pour valeur $y - x$ si $y - x > 0$ et $1 + y - x$ si $y - x < 0$. Une telle variable aléatoire est bien comme on peut le vérifier détail, indépendante de H.

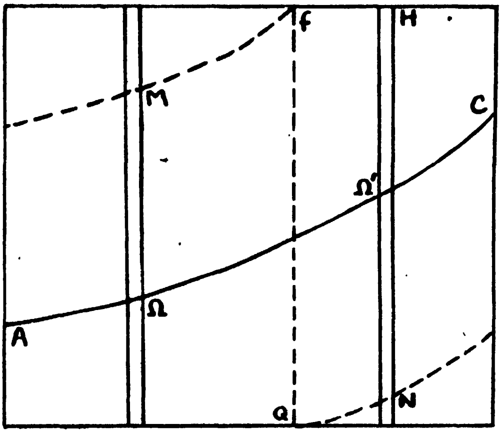


Fig. 2

On voit qu'il existe une infinité de familles associées à une famille donnée, mais les courbes représentatives auront dans la représentation primitive des points à l'infini, ou en tout cas, seront composées de morceaux. C'est ce qui donne à la solution particulière $\varphi = F^{x_1}(x_2)$ une importance toute spéciale.

RÉSULTATS GÉNÉRAUX

Il est bien clair que ces résultats s'étendent à un espace à un nombre fini quelconque de dimensions (et même à une infinité dénombrable). De sorte qu'on peut, d'une infinité de manières, représenter une loi de probabilité quelconque (absolument continue) par des expressions de la forme:

$$x_i = f_i(\alpha_1, \alpha_2, \ldots, \alpha_n)$$

$i = 1, 2, \ldots, n$, où les α_i peuvent être supposées ramenées à être uniformes de zéro à un, et sont des variables indépendantes.

Comme d'autre part, toute distribution à deux, trois, ..., variables peut être considérée comme projection d'une distribution à un nombre supérieur quelconque, de dimensions, on voit que toute loi, à deux variables, par exemple, peut être d'une infinité de manières représentée avec un nombre quelconque de facteurs indépendants.

UN CAS PARTICULIEREMENT INTÉRESSANT

On sait que les théories analogues à celle de Spearman, emploient des facteurs communs et facteurs de groupes pour la représentation des corrélations.

Conçue d'une manière générale, cette idée prend la forme suivante: Soient n variables $x_1, x_2, \ldots, x_n$, et N facteurs $\alpha_1, \alpha_2, \ldots, \alpha_n$. Considérons un groupe Γ de facteurs qui apparaîtra dans tous les x_1, et n groupes $G_1, G_2, \ldots, G_n$, sans partie commune. Les expressions: $x_i = f_i(\Gamma, G_i)$ fournissent une généralisation des schemas linéaires classiques. Il est à remarquer qu'on peut, de façon très générale, obtenir une forme simple de la fonction caractéristique de la loi des x_i. En effet: $\phi(u_1, u_2, \ldots, u_n) = Ee^i(u_1x_1+u_2x_2+\ldots+u_nx_n)$ peut être obtenue en deux étapes. Laissant d'abord constants les facteurs du groupe Γ, les x_i ainsi liés sont indépendants entre eux, de sorte que la première intégration donne une expression de la forme:

$$\phi_1(\Gamma, u_1) \quad \phi_2(\Gamma, u_2) \quad \ldots \quad \phi_n(\Gamma, u_n)$$

et que par une dernière intégration, on a la forme explicite générale

$$\phi(u_1\ u_2, \ldots, u_n) = \int \phi_1(\Gamma, u_1) \ldots \phi_n(\Gamma, u_n) d\Gamma$$

en notant $d\Gamma$ l'élement de probabilité du groupe des facteurs communs. Sans étudier plus avant la question de savoir comment on reconnaitra si les x_i sont effectivement liés en probabilité et de quelle manière ils le sont, on doit remarquer qu'on peut trouver des différences assez profondes avec le schéma linéaire avec des coéfficients de corrélation nuls, ce qui ne saurait se produire dans le cas linéaire avec un seul facteur commun.

UN CAS PARTICULIER

Soit x uniforme de 0 à 1, α uniforme de $-1/2$ à $+1/2$, et indépendant de x. La relation $y = \alpha x$ définit dans un domaine traiangulaire une loi de probabilité à deux variables dont la loi élémentaire est donnée par: $dx.dy/x$. Il y a évidemment liaison de probabilité, la loi de y liéee est uniforme mais d'étendue proportionnelle à x. Il est clair que la moyenne de y liée est nulle, ainsi que le moment $E[y(x-1/2)]$ ce qui donne ce qu'on peut appeler l'indépendance au second ordre.

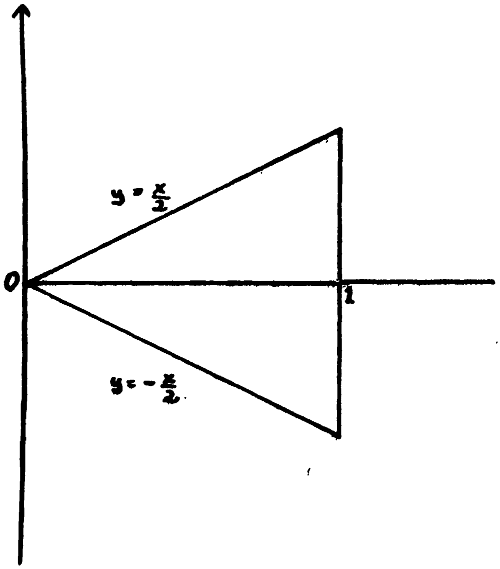


Fig. 3

LES PROBLEMES PRATIQUES

On peut en se bornant au cas le plus simple du facteur associé à la variable x_1, indiquer quelques propriétés élémentaires. Reprenant les notations x et y et notant α le facteur associé à x, on aura $y = f(x,\alpha)$ où α peut être supposé à répartition uniforme de 0 à 1.

On voit que cette forme de relation met tout à fait en évidence des éléments importants. Si l'on fait $\alpha = 1/2$, on a ce qu'on pourrait appeler la ligne médiane (ou ligne de régression médiane, si ce mot de régression ne parait pas déjà étendu à l'excès). On aurait de même les lignes quartiles, etc.

On peut se demander en particulier si toutes ces lignes pourraient être des droites. La surface déjà considérée: $\alpha = F^x(y)$ serait alors une surface réglée. D'autre part, f doit être monotone en α. Il faudrait que l'on eût:

$$y = xp(\alpha) + q(\alpha)$$

Si le point xy peut parcourir toute la génératrice, le plan tangent fait un tour complet à moins que la surface ne soit développable. Il faut donc, si le point parcourt toute la génératrice que la surface soit un cylindre de génératrices parallèles au plan directeur, soit $p(\alpha) = C^{te} = r_0$, et $q(\alpha)$ monotone. Ainsi toutes les droites sont parallèles.

Puis il peut aussi arriver que x soit borné. On peut le supposer positif, et la condition de monotonie ne présente plus aucune difficulté. Le lecteur construira aisément de nombreux exemples, nous venons d'en donner plus haut un très simple.

PROBLEMES D'AJUSTEMENT

On peut remarquer que, dans un cas où la ligne médiane est supposée droite, le problème de la recherche de cette droite est analogue à celui de la droite de régression ordinaire. Il convient alors d'insister sur le fait que la méthode des moindres carrés devrait être replacée par la méthode des moindres distances. Il faudrait rechercher une droite telle que la some des distances parallèlement à l'axe des y (somme pondérée s'il y a lieu), soit un minimum. On trouve alors aisément

1) que la droite cherchée doit être une droite médiane dans le plan, c'est-à-dire que les masses qu'elle sépare doivent être ou égales, ou vérifier les conditions habituelles.

2) que la droit qui joint les centres de gravité des masses égales doit être parallèle à l'axe des y (j'énonce cette condition dans le cas le plus simple).

PROBLEMES D'OPTIMUM

Beaucoup plus encore que dans l'analyse factorielle classique, il faut, pour choisir dans la famille infinie des solutions possibles, imposer des conditions de simplicité et d'optimum adaptées à question que l'on étudie.

Il est certain que dans une loi de Laplace-Gauss à deux variables, avec une surface en cloche qui soit allongée dans un certain sens, l'un des facteurs peut être rattaché à des parallèles ayant cette direction, l'autre facteur étant rattaché aux droites associées, ici conjuguées. Le premier facteur variera très rapidement, et on peut lui imposer une condition d'extremum. S'il s'agit au contraire d'une cloche très pointue, mais à peu près ronde, il serait naturel d'envisager un facteur rattaché à des courbes fermées partant du sommet, et un facteur d'orientation des courbes associées.

Si l'on imposait par exemple aux courbes fermées de renfermer l'aire la plus petite pour une probabilité donnée, on trouverait les ellipses de la loi et on pourrait prendre comme associées des droites rayonnant autour du centre.

UN EXEMPLE ASSEZ SIMPLE DE LIAISON PAR FACTEUR COMMUN

On peut étudier en général le cas où l'on aurait:

$$x_1 = a_1 \alpha_1 + c_1 \alpha_3 + d_1 \alpha_1 \alpha_3$$
$$x_2 = a_2 \alpha_2 + c_2 \alpha_3 + d_2 \alpha_2 \alpha_3$$

Les a, c, d sont des constantes. On peut former la fonction caractéristique $Ee^{i}\,(u_1x_1+u_2x_2)$ par la méthode générale indiquée.

Bornons-nous au cas très simple où l'on a:

$$x_1 = \alpha_1 \alpha_3$$
$$x_2 = \alpha_2 \alpha_3$$

Les α_1, α_2, α_3 suivent indépendamment une loi de Laplace Gauss centrée et normée. On obtient aisément la fonction caractéristique:

$$\phi\,(u_1\,u_2) = \frac{1}{\sqrt{1 + u_1^2 + u_2^2}}$$

On se trouve encore dans un cas où le coéfficient de corrélation est nul. La loi obtenue d'ailleurs évidemment isotrope dans le plan $x_1\,x_2$, et l'on peut remarquer que, dans ce plan, les cercles:

$$x_1^2 + x_2^2 = \beta_1^2$$

et les droites issues de l'origine $x^2/x^1 = tg\beta_2$ forment un réseau de courbes associées. On a donc une autre représentation avec deux facteurs.

CONCLUSIONS

Les propriétés très simples que nous avons indiquées et qu'on pourrait aisément développer paraîtront peut-être, au moins pour certaines, plus curieuses qu'utiles. Il me semble qu'elles éclairent le problème général de l'analyse factorielle en montrant quelle large indétermination il présente dès qu'on sort du domaine, déjà très étendu, des schémas linéaires. On voit en tout cas qu'on peut espérer représenter, par ces méthodes, des liaisons de probabilité qui échappent aux méthodes classiques d'analyse.

Résumé

1) It is possible to generalize classical factor analysis by still considering independent factors, but permitting the variates studied to be more general functions of these factors than linear functions.

2) In the case of an absolutely continuous bivariate probability distribution, it is always possible, and such in an infinite number of ways, to represent it with two independent factors, one of which can be chosen arbitrarily.

For a trivariate distribution it is possible to associate to any congruence of curves in space three independent variables, one of them associated with the curves of the congruence, the other two being two independent variables which define the congruence. Those results apply to any number of variates, and permit representation of a distribution with any number of variates with the aid of an equal or any larger number of independent factors. It is in particular possible to generalize Spearman's theory with common factors.

3) The special case $y = f(x,\alpha)$ where the two independent factors are x and α, gives evidence of simple and useful elements such as the locus of the representative point of the median value of y conditioned by x.

4) Linear factor analysis already raises problems with regard to adaptation to an observed sample by conditions of simplicity and optimum. These problems are the same in the generalized case, but the choice must be made among a much wider class of possible solutions. It is at any rate possible to represent by means of these methods, probability distributions where the variables, although dependent, are not correlated.

5) The simplest generalization of linear factor analysis seems to be the bilinear case. A simple example is given for bivariate distributions.

UNE NOUVELLE METHODE D'ANALYSE FACTORIELLE

par Pierre Delaporte

Professeur a l'Institut de Statistique de l'Université de Paris (France)

Les divers caractères biométriques, physiologiques, psychologiques que l'on peut mesurer sur un même individu sont statistiquement liés et de nombreux problèmes d'analyse des corrélations qui existent entre ces caractères ont été proposés. L'un des plus intéressants parmi ces problèmes, mais peut être aussi le plus difficile de tous, consiste à rechercher si les liaisons que l'on observe entre les divers caractères étudiés pourraient provenir d'une origine unique essentiellement liée à l'individu et agissant sur tous les caractères que l'on peut mesurer sur cet individu.

Dès 1904, C. Spearman [11],[1] en examinant les coefficients de corrélation obtenus entre plusieurs tests psychologiques, observait une hiérarchie entre ces coefficients et, par des études successives, était conduit à penser que tous ces résultats pourraient provenir d'une unique source de liaison, *le facteur général* commun à tous les caractères que l'on peut observer sur un même individu.

Le résultat de la mesure d'un caractère est ainsi formé d'une partie provenant du facteur général et d'une autre partie propre au caractère choisi et aux erreurs commises de cette mesure. Spearman a désigné cette seconde partie sous le nom de *facteur spécifique* du caractère. Les facteurs spécifiques seront donc par hypothèse statistiquement indépendants du facteur général. Les facteurs spécifiques seront statistiquement indépendants les uns des autres s'il s'agit de caractères ou d'aptitudes entièrement distincts les uns des autres.

En fait il est généralement difficile d'avoir des caractères entièrement distincts les uns des autres, par exemple deux tests cherchant à mesurer des aptitudes différentes ne pourront éviter complètement d'avoir, en plus du facteur général, une partie commune telle que la mémoire ou la rapidité des réactions. Il en résulte que les tests ou caractères choisis ne pourront se conformer entièrement aux hypothèses initiales de C. Spearman. L'ensemble des liaisons que l'on observe sur plusieurs caractères nécessitera l'adjonction au facteur général d'un ou de plusieurs *facteurs de groupes*, facteurs qui se superposent au facteur général mais qui s'adressent seulement à quelques uns des caractères ou aptitudes mesurés.

---

[1] Les numéros entre crochets renvoient à la bibliographie.

L. L. Thurstone a donné une importance bien plus grande à ces facteurs de groupes et leur a attribué un rôle analogue à celui du facteur général de Spearman; pour Thurstone cependant il n'existe pas nécessairement un facteur général mais seulement une superposition de facteurs, appelés *facteurs multiples*. Le résultat de chaque test est ainsi formé d'une partie imputable à plusieurs facteurs multiples et d'une autre partie, independante des facteurs multiples, spécifique du test et de l'erreur commise dans la mesure.

Ces deux théories admettent en fait l'existence de plusieurs facteurs indépendants les uns des autres.

Nous adopterons les hypothèses suivantes:

1°—la mesure de chaque caractère peut s'exprimer par une fonction linéaire du facteur général, de facteurs de groupes s'il y a lieu, et du facteur spécifique du caractère.

2°—facteur général, facteurs de groupes et facteurs spécifiques sont tous statistiquement indépendants entre eux.

Je vais montrer comment, en introduisant la Statistique Mathématique en Analyse Factorielle, il m'a été possible de mettre au point une méthode permettant de:

1°—trouver, d'après les observations, le nombre de facteurs de groupes et les caractères dans lesqüels chacun d'eux entre.

2°—estimer statistiquement les coefficients de corrélation entre le facteur général, les facteurs de groupes et chacun des tests.

Juger si les valeurs obtenues sont statistiquement significatives.

Supposons que l'on mesure k caractères $1, 2, \ldots, i, \ldots, k$ sur chacun des individus.

La valeur ${}_jX_i$ du caractère i sur l'individu j peut être exprimée sous la forme de variable normée ${}_jx_i = \dfrac{{}_jX_i - E[X_i]}{\sigma_{x_i}}$,

où $E[X_i]$ est l'espérance mathématique de la valeur du caractère i dans la population σ_{x_i} est l'écart type des valeurs du caractère i dans la population.

Désignons par:

${}_jG$ *le facteur général* de l'individu j exprimé en variable normée,

${}_jB$, ${}_jC$, ... *les facteurs de groupes* de l'individu j exprimé en variables normées,

${}_jS_i$ *le facteur spécifique* de l'individu j pour le caractère i, exprimé en variable normée.

La première hypothèse donne des relations de la forme suivante pour chaque individu j:

$$(1) \qquad \begin{cases} {}_jx_1 = \alpha_1\,{}_jG + \beta_1\,{}_jB + \gamma_1\,{}_jC + \dots + \lambda_1\,{}_jS_1 \\ \dots \qquad \dots \qquad \dots \\ {}_jx_i = \alpha_i\,{}_jG + \beta_i\,{}_jB + \gamma_i\,{}_jC + \dots + \lambda_i\,{}_jS_i \\ \dots \qquad \dots \qquad \dots \\ {}_jx_k = \alpha_k\,{}_jG + \beta_k\,{}_jB + \gamma_k\,{}_jC + \dots + \lambda_k\,{}_jS_k \end{cases}$$

avec

$$E[x_1] = \dots = E[x_i] = \dots = E[x_k] = 0$$

$$E[G] = E[B] = E[C] = \dots = E[S_1] = \dots = E[S_i] = \dots = E[S_k] = 0$$

$$\sigma_{x_1} = \dots = \sigma_{x_i} = \dots = \sigma_{x_k} = \sigma_G = \sigma_B = \sigma_C = \dots = \sigma_{S_1} = \sigma_{S_i} = \dots = \sigma_{S_k} = 1$$

Pour tout caractère i ne nécessitant pas l'introduction des facteurs de groupes B ou C ou ... les coefficients β_i ou γ_i ou ... respectivement seront nuls par hypothèse. Les autres coefficients α, β, γ, ..., λ sont des constantes à déterminer pour chacun des caractères.

En désignant par ρ le coefficient de corrélation totale de la population, l'hypothèse 2, sur l'indépendance des facteurs entre eux, peut s'écrire ainsi

$$\rho_{GB} = \rho_{GC} = \dots = \rho_{GS_i} = \rho_{BS_i} = \rho_{CS_i} = \dots = 0 \qquad (i = 1, 2, \dots, k)$$

$$\rho_{S_hS_i} = 0 \qquad (h \neq i;\ h, i = 1, 2, \dots, k)$$

Le coefficient de corrélation vrai entre 2 caractères h et i est donné par:

$$(2) \qquad \rho_{hi} = E[x_h \cdot x_i] = \alpha_h\,\alpha_i + \beta_h\,\beta_i + \gamma_h\,\gamma_i + \dots$$

Dans le cas particulier où il n'existe pas de facteur de groupe, Spearman a établi une des conditions nécessaires pour qu'un ensemble de k caractères soit représentable par le schéma 1 facteur général et k facteurs spécifiques en se limitant aux moments-produits du premier ordre.

En remarquant que:

$$\rho_{hi} = E[x_h\,x_i] = \alpha_h\,\alpha_i$$

on en déduit que les coefficients de corrélation ρ doivent vérifier la condition dite de nullité des Tétrades:

$$\tau_{ghij} = \rho_{gh}\,\rho_{ij} - \rho_{gi}\,\rho_{hj} = 0 \qquad (g \neq h \neq i \neq j;\ g, h, i, j = 1, \dots, k)$$

Cette condition nécessaire est celle que doivent vérifier les moments-produits du premier ordre à 2 variables. Il est évident que si l'on utilisait des moments-produits d'ordres supérieurs on obtiendrait d'autres conditions nécessaires. Garnett a montré [9] que si l'ensemble des k caractères était distribué selon une loi de Laplace-Gauss à k variables, la nullité de l'ensemble des tétrades était la condition necessaire et suffisante pour que l'ensemble des caractères puisse être représenté par un facteur général et k facteurs spécifiques independants.

Enfin G. Darmois [1 et 2] en utilisant la fonction caractéristique de l'ensemble des variables, a montré que si les k caractères étaient distribues selon une loi de Laplace-Gauss à k variables la décomposition en 1 facteur général et k facteurs spécifiques était unique.

Pour conserver les deux propriétés données par Garnett et Darmois il faudra donc transformer la distribution à étudier en distribution gaussienne à k variables; ce qu'on obtiendra pratiquement en transformant en distribution gaussienne chacune des distributions des caractères qui s'en éloignerait fortement.

1. LE PROBLEME STATISTIQUE

En fait le problème à résoudre n'est pas de Calcul de Probabilités mais de Statistique: on ne connait pas l'ensemble de la population mais seulement un échantillon de n individus supposé prélevé au hasard dans une population inconnue. En outre on ne dispose pas des valeurs réelles des caractères, mais seulement de mesures entachées d'erreurs.

Désignons par r les coefficients de corrélation empiriques. Formons les tétrades empiriques

$$t_{ghij} = r_{gh}\, r_{ij} - r_{gi}\, r_{hj} \qquad (g \neq h \neq i \neq j;\ g, h, i, j = 1, \ldots, k)$$

Les tétrades t ne sont pas numériquement nulles et on doit chercher si la valeur de t est statistiquement compatible avec l'hypothèse $\tau = 0$, la valeur observée étant alors attribuée aux erreurs d'échantillonnage.

Bien que K. Pearson [10] ait montré que la loi de distribution de t est théoriquement et pratiquement différente de la loi de Laplace-Gauss, on forme habituellement le rapport t/σ_t et on le compare à la loi de Laplace-Gauss. Ce procédé est très long car pour 20 variables, par exemple, il nécessite le calcul de 14,535 tétrades et écart types σ_t; le jugement du résultat reste difficile au point de vue statistique.

Si certaines tétrades empiriques ne peuvent être statistiquement considérées comme nulles, le schéma de Spearman n'est pas applicable et il faut introduire un ou plusieurs facteurs de groupes. La recherche des caractères qui nécessitent un facteur de groupe pour représenter les corrélations observées est pratiquement impossible à faire d'après les tétrades.

2. NOUVELLE MÉTHODE DE RECHERCHE DE LA POSITION DES FACTEURS DE GROUPES

D'après l'hypothèse de l'existence d'un facteur général, sans facteur de groupe, le critère de nullité des tétrades

$$\tau_{gjih} = 0 \quad \text{peut s'écrire} \quad \rho_{gj}\, \rho_{hi} = \rho_{gi}\, \rho_{hj}$$

ou encore

$$\frac{\rho_{gj}}{\rho_{hj}} = \frac{\rho_{gi}}{\rho_{hi}} = \frac{\alpha_g \alpha_j}{\alpha_h \alpha_j} = \frac{\alpha_g \alpha_i}{\alpha_h \alpha_i} = \frac{\alpha_g}{\alpha_h}$$

d'une manière générale

$$(3)\ \frac{\rho_{g1}}{\rho_{h1}} = \frac{\rho_{g2}}{\rho_{h2}} = \ldots = \frac{\rho_{gi}}{\rho_{hi}} = \frac{\rho_{gj}}{\rho_{hj}} = \ldots = \frac{\rho_{gk}}{\rho_{hk}} = \frac{\alpha_g}{\alpha_h} \qquad (i, j \neq g, h)^2$$

ainsi une seule suite d'égalités résume C^2_{k-2} tétrades et fait intervenir $2k-4$ coefficients de corrélation différents.

Si l'on forme les $k-1$ relations du type précédent construites sur la base du caractère h et égales à:

$$\frac{\alpha_1}{\alpha_h}, \frac{\alpha_2}{\alpha_h}, \ldots, \frac{\alpha_g}{\alpha_h}, \ldots, \frac{\alpha_k}{\alpha_h}$$

on fait intervenir au moins 2 fois chacun des coefficients de corrélation ρ.

Si l'on remplace les coefficients de corrélation ρ de la population par les coefficients de corrélation r observés sur un échantillon de n individus les quotients de coefficients de corrélation ne vérifient plus exactement les égalités (3).

Il faut donc rechercher si, en tenant compte des erreurs d'échantillonnage, les quotients formés des r empiriques vérifient statistiquement la relation (3) avec une probabilité P qu'on se donnera à l'avance.

3. LOI DE DISTRIBUTION DES RAPPORTS

Si les caractères étudiés sont distribués selon une loi de Laplace-Gauss, le nombre n des individus mesurés doit être nécessairement grand (plusieurs centaines) pour que l'analyse factorielle puisse donner des résultats statistiquement intéressants.[3] Si le coefficient de corrélation de la population ρ n'est voisin ni de $+1$, ni de -1, r est approximativement distribué selon une loi de Laplace-Gauss telle que:

$$E[r] = \rho\left(1 - \frac{1-\rho^2}{n} + \ldots\right)$$

$$\sigma_r = \frac{1-\rho^2}{\sqrt{n}}$$

On obtiendrait une approximation meilleure [4] valable jusqu'à $\rho = \pm 1$ et pour des nombres n petits, en utilisant les covariances γ_{gh} au lieu des

---

[2] Cette relation correspond à la proportionalité de 2 colonnes du tableau des coefficients de corrélation qu'avait observée C. Spearman dans ses premiers travaux.

[3] Voir au paragraphe 6 l'erreur type des estimations des coefficients $\alpha, \beta, \gamma, \ldots$

coefficients de corrélation ρ_{gh} qui remplacent la relation (3) des rapports des coefficients de corrélation par

$$(4)\qquad \frac{\gamma_{g1}}{\gamma_{h1}} = \frac{\gamma_{g2}}{\gamma_{h2}} = \ldots = \frac{\gamma_{gi}}{\gamma_{hi}} = \frac{\gamma_{gj}}{\gamma_{hj}} = \ldots = \frac{\gamma_{gk}}{\gamma_{hk}} = \frac{\alpha_g\,\sigma_{x_g}}{\alpha_h\,\sigma_{x_h}} \qquad (i, j \neq g, h)$$

On peut montrer que la loi de distribution d'un rapport $\Phi = C_{gi}/C_{hi}$ des covariances empiriques est donné par

$$\rho(\phi)d\phi = \frac{\sigma_g\,\sigma_h\,\sqrt{1-\rho^2}}{\pi} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{\gamma_g^{\,2}}{\sigma_g^{\,2}} + 2\rho\frac{\gamma_g\,\gamma_h}{\sigma_g\,\sigma_h} + \frac{\gamma_h^{\,2}}{\sigma_h^{\,2}}\right)} \frac{d\phi}{\phi^2\sigma_h^{\,2} - 2\rho\phi\sigma_g\sigma_h + \sigma_g^{\,2}}$$

$$+ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\gamma_h^{\,2}(\phi-\varphi)^2}{\phi^2\sigma_h^{\,2} - 2\rho\phi\sigma_g\sigma_h + \sigma_g^{\,2}}} \cdot \frac{\gamma_h\sigma_g\sigma_h\,[-\phi\varphi\sigma_h^{\,2} + \rho(\phi+\varphi)\,\sigma_g\sigma_h - \sigma_g^{\,2}]}{(\phi^2\sigma_h^{\,2} - 2\rho\phi\sigma_g\sigma_h + \sigma_g^{\,2})^{3/2}}$$

$$\times \left[\int_0^{\frac{\gamma_h}{\sigma_g\sigma_h}\frac{-\phi\varphi\sigma_h^{\,2} + \rho(\phi+\varphi)\sigma_g\sigma_h - \sigma_g^{\,2}}{\phi^2\sigma_h^{\,2} - 2\rho\phi\sigma_g\sigma_h + \sigma_g^{\,2}}} \frac{2}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} d\lambda\right] d\phi$$

où γ_g et γ_h représentent respectivement les covariances γ_{gi} et γ_{hi}, σ_g et σ_h désignent les erreurs types[4] des covariances:

$$\sigma_g = \sigma_{x_g}\sigma_{x_j}\sqrt{(1+\rho_{gi})/n} \qquad\qquad \sigma_h = \sigma_{x_g}\sigma_{x_h}\sqrt{(1+\rho_{hi})/n}$$

$$\rho = \rho_{c_{gi}\,c_{hi}} = (\rho_{gh} + \rho_{gi}\,\rho_{hi})/\sqrt{(1+\rho^2_{gi})(1+\rho^2_{hi})} \qquad \text{et } \varphi = \gamma_{gi}/\gamma_{hi}$$

Les courbes représentatives de la densité de probabilité ont des formes très variées selon les valeurs des paramètres:

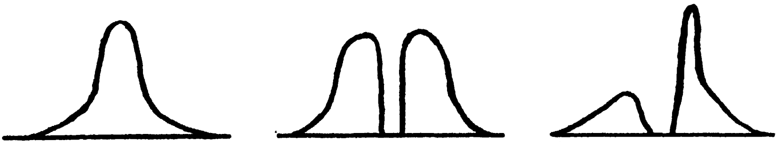


Fig. 1

Cette densité de probabilité permet de déterminer les limites entre lesquelles Φ doit se trouver avec une quasi certitude P de jugement donnée à l'avance.

---

[4] Erreur type = écart type de l'erreur d'échantillonnage,

Un cas particulier très fréquent dans les applications est celui où

$$\frac{\gamma_h}{\sigma_h} = \frac{(\rho_{hi}\,\sigma_{x_h}\,\sigma_{x_i})}{\sqrt{(1+\rho^2_{hi})/n}\;\sigma_{x_h}\,\sigma_{x_i}} = \sqrt{n}\,\frac{\rho_{hi}}{\sqrt{1+\rho^2_{hi}}}$$

est grand, alors le premier terme de la densité de probabilité est très voisin de zéro, l'intégrale qui figure dans le deuxième terme est très voisine de 1 et, en posant:

$$W = \frac{\gamma_h\,(\phi - \varphi)}{\sqrt{\phi^2\,\sigma^2_h - 2\rho\phi\sigma_g\sigma_h + \sigma^2_g}}$$

la probabilité élémentaire peut se mettre sous la forme

$$f(\phi)\,d\phi \simeq \frac{1}{\sqrt{2\pi}}\,e^{-\frac{W^2}{2}}\,dW$$

donc, si à la quasi certitude P il correspond un écart w d'après la loi de Laplace-Gauss, les limites de confiance de ϕ sont données par

$$\lim \phi = \frac{\rho\sigma_g\sigma_h w^2 - \varphi\gamma^2_h \pm \sqrt{(\rho\sigma_g\sigma_h w^2 - \varphi\gamma^2_h)^2 - (w^2\sigma^2_h - \gamma^2_h)(w^2\sigma^2_g - \gamma^2_h\varphi^2)}}{w^2\,\sigma^2_h - \gamma^2_h}$$

Ainsi, chacun des rapports C_{gi}/C_{hi} permet de déterminer l'intervalle dans lequel doit se trouver, avec la probabilité P, la quantité à estimer $\alpha_g\sigma_{x_g}/(\alpha_h\sigma_{x_h})$. Si tous ces intervalles ont une partie commune, la relation d'égalité des rapports des covariances (4) est statistiquement vérifiée, et si toutes les relations analogues obtenues en faisant varier g et h le sont aussi, toutes les tétrades sont statistiquement nulles. On peut alors admettre que l'ensemble des corrélations qui existent entre les caractères 1, 2, ..., k est explicable par la seule présence d'un facteur général.

4. APPROXIMATION DE LA LOI DE DISTRIBUTION DES RAPPORTS POUR LES APPLICATIONS PRATIQUES [5]

Dans les applications à des données expérimentales, j'ai cherché à supprimer tout calcul numérique pour déterminer les intervalles communs ci-dessus. Pour limiter la dimension des abaques à construire, j'ai utilisé l'égalité (3) des rapports de coefficients de corrélation au lieu de celle de covariances.

Examinons successivement ce que devient l'égalité (3) pour des mesures entachées d'erreurs, comment construire les abaques et les graphiques d'intervalles communs.

4–1. *Effet des erreurs de mesure sur le critère*

Si les mesures des caractères sont entachées d'erreurs indépendantes entre elles et indépendantes de la grandeur du caractère à mesurer, on peut montrer que les suites d'égalités (3) des rapports de coefficients de corrélation existent encore pour les coefficients atténués.

4–2. *Construction des abaques donnant les intervalles d'échantillonnage des rapports r_{gi}/r_{hi}*

Si nous négligeons la corrélation qui existe entre $(r_{gi}-\rho_{gi})$ et $(r_{hi}-\rho_{hi})$, on peut, pour un nombre donné n d'individus de l'échantillon, former un abaque donnant les limites de l'intervalle d'échantillonnage que peut atteindre le rapport r_{gi}/r_{hi} lorsque chacun des coefficients de corrélation empirique prend toutes les valeurs possibles autour de ρ_{gi} et ρ_{hi} avec une quasi certitude donnée P.

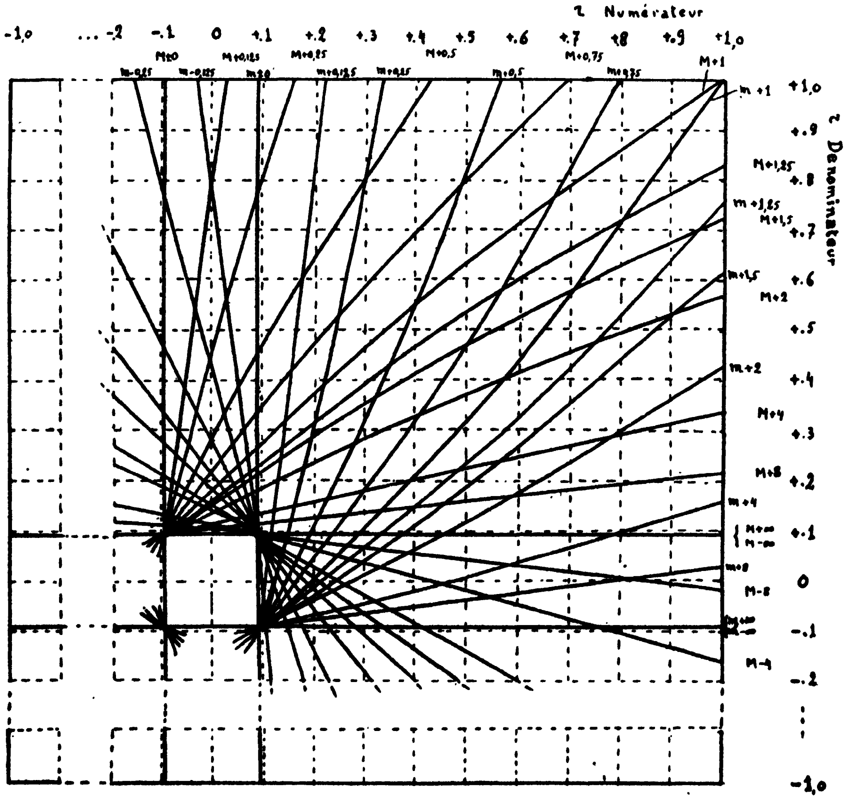


Fig. 2. Forme schématique d'un abaque pour $n = 300$

Pour former nos abaques, nous avons utilisé la quasi-certitude $P = 0.90$, c'est-à-dire que si $p(r/n, \rho)$ indique la densité de probabilité d'obtenir un coefficient de corrélation r dans un échantillon de n individus prélevés au hasard dans une population distribuée suivant une loi de Laplace-Gauss à 2 variables de coefficient de corrélation ρ on α:

$$(5) \qquad \int_{-1}^{l^r} p(r\,|\,n, \rho)\, dr = 0.05 \quad \text{et} \quad \int_{L^r}^{+1} p(r\,|\,n, \rho)\, dr = 0.05$$

Les abaques construits pour $n = 100$ et $n = 200$ ont été obtenus au moyen des tables de F. N. David [3]. Les abaques relatifs à $n = 300$, $n = 500$ et $n = 1000$ ont été obtenus au moyen de la transformation de R. A. Fisher qui donne pour valeurs limites l^r et L^r de r

$$(6) \quad L^r = \operatorname{tgh}\left(\arg\operatorname{tgh} \rho + \frac{v}{\sqrt{n-3}}\right) \text{ et } l^r = \operatorname{tgh}\left(\arg\operatorname{tgh} \rho - \frac{v}{\sqrt{n-3}}\right)$$

v étant un écart tel que, d'après la loi de Laplace-Gauss réduite, la probabilité d'observer un écart inférieur à v soit la quasi-certitude utilisée $P = 0.9$.

Des relations (5) ou (6) on tire la forme réciproque donnant les valeurs limites de ρ acceptables pour une valeur observée r et on construit l'abaque donnant les limites entre lesquelles doit se trouver α_g/α_h d'après la valeur de r_{gi}/r_{hi} observée.

Les intervalles lus sur l'abaque sont tels que, d'après la relation (2) chacun des intervalles relatifs à r_{gi}/r_{hi} a une probabilité comprise entre 0.01 et 0.1 de contenir α_g/α_h, cette probabilité étant 0.01 si $\rho_{r_{gi}\,r_{hi}} = 0$ et 0.1 si $\rho_{r_{gi}\,r_{hi}} = 1$.

4-3. *Les graphiques d'intervalles communs*

Formons un graphique en portant en ordonnée les valeurs de ρ_{gi}/ρ_{hi} lues sur l'abaque, qui varient de $-\infty$ à $+\infty$, avec une échelle permettant une représentation pratique de l'ensemble des valeurs et en prenant en abscisse successivement les divers caractères ($i = 1, 2, \dots, k$; $i \neq g, h$).

Pour chaque i on trace le ou les intervalles entre lesquels doit se trouver le rapport ρ_{gi}/ρ_{hi} et on cherche s'ils ont une partie commune ou quels sont les caractères pour lesquels il faut éliminer les rapports r_{gd}/r_{hd} pour que les autres aient une partie commune. Cette partie commune (figurée par des lignes en pointillé) limite le domaine dans lequel α_g/α_h doit se trouver.

On forme, les uns à côté des autres, la série des $k-1$ diagrammes analogues en h estimant

$$\frac{\alpha_1}{\alpha_h},\ \frac{\alpha_2}{\alpha_h},\ \frac{\alpha_3}{\alpha_h},\ \ldots,\ \frac{\alpha_i}{\alpha_h},\ \ldots,\ \frac{\alpha_k}{\alpha_h}$$

puis la série des $k-1$ diagrammes relatifs à un autre caractère j

$$\frac{\alpha_1}{\alpha_j},\ \frac{\alpha_2}{\alpha_j},\ \frac{\alpha_3}{\alpha_j},\ \ldots,\ \frac{\alpha_i}{\alpha_j},\ \ldots,\ \frac{\alpha_k}{\alpha_j}$$

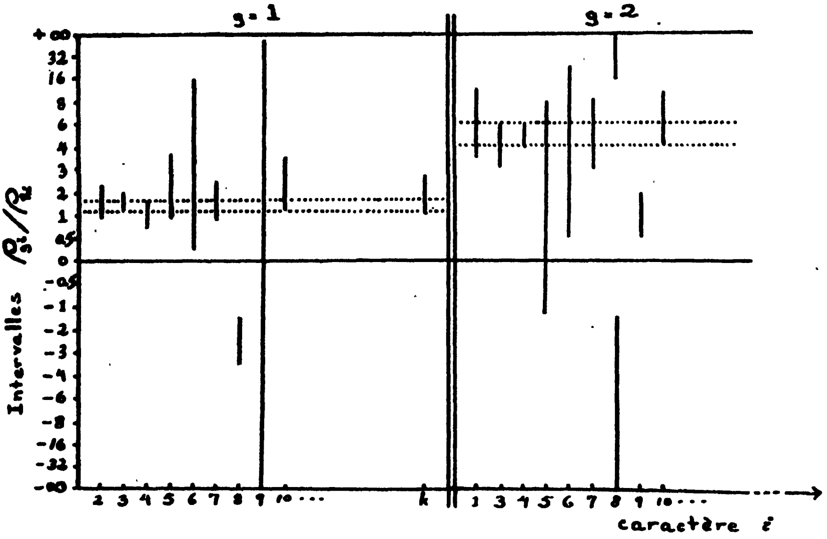


Fig. 3

Par l'ensemble de ces diagrammes, on recherche chacun des coefficients de corrélation qu'il faut éliminer pour que l'ensemble des autres donne des intervalles ayant toujours une partie commune et soit donc représentable par un facteur général; les autres nécessiteront l'introduction de facteurs de groupes.

Remarquons que s'il existe un intervalle commun à tous les $k-2$ caractères de chaque diagramme, la nullité de l'ensemble des tétrades est statistiquement vérifiée et les k caractères sont représentables par le schéma de Spearman avec un facteur général et des facteurs spécifiques, sans facteur de groupe. Ainsi pour 20 caractères qui fournissaient 14,535 tétrades à calculer et dont il fallait examiner statistiquement la nullité, il suffit, par ce nouveau procédé, de lire sur l'abaque les limites de confiance de 684 rapports et d'en construite le graphique pour résoudre le problème.

D'autre part, si l'on constate qu'une tétrade ne peut statistiquement être considérée comme nulle, il faut rechercher parmi les 4 caractères qu'elle introduit lequel ou lesquels doivent être éliminés. Au contraire le procédé des intervalles communs que nous proposons ne peut laisser d'hésitation que sur 3 caractères et le plus souvent le caractère à éliminer est celui qui intervient simultanément au numérateur et au dénominateur.

La simplification ainsi apportée permet de faire un dépouillement sous forme de tableau à double entrée des coefficients de corrélation pour lesquels il y a doute et de déterminer ainsi la position et le nombre des facteurs de groupes sans aucun calcul préalable. Les autres méthodes de localisation des facteurs de groupes proposées jusqu'à présent ne peuvent atteindre ce résultat que par des approximations successives démontrée.

5. ESTIMATION DU COEFFICIENT DE CORRÉLATION ENTRE UN CARACTERE ET LE FACTEUR GÉNÉRAL

D'après les hypothèses de Spearman, les coefficients $\alpha_1, \dots, \alpha_k$ du système (1) sont les coefficients de corrélation $\rho_{1G}, \dots, \rho_{kG}$ entre les caractères 1, ..., k et le facteur général G. On peut les obtenir au moyen de la relation

$$\rho^2_{hG} = \frac{\rho_{1h}\,\rho_{2h}}{\rho_{12}} = \frac{\rho_{1h}\,\rho_{3h}}{\rho_{13}} = \dots = \frac{\rho_{(k-1)h}\,\rho_{kh}}{\rho_{(k-1)k}}. \tag{7}$$

Les divers termes de cette relation ne sont plus égaux lorsqu'on remplace les coefficients de corrélation théoriques ρ par leurs valeurs r observées sur l'échantillon. Pour obtenir une meilleure estimation que celle donnée par un seul des termes de la suite d'égalités (7), Spearman [12] a proposé de prendre la somme des numérateurs divisée par la somme des dénominateurs:

$$\rho^2_{hG} = \frac{\rho_{1h}\,\rho_{2h} + \rho_{1h}\,\rho_{3h} + \dots + \rho_{(k-1)h}\,\rho_{kh}}{\rho_{12} + \rho_{13} + \dots + \rho_{(k-1)k}}$$

L'estimation r_{hG} de ρ_{hG} que l'on obtient en remplaçant les coefficients de corrélation ρ par leurs valeurs observées r ne tient pas compte de la loi de probabilité de l'erreur d'échantillonnage des coefficients de corrélation observés. On peut obtenir des resultats beaucoup plus précis en prenant pour estimation de ρ^2_{hG} la moyenne pondérée des divers termes de la relation (7) *ayant l'écart type minimum.* Si l'on désigne par $\sigma_{h\cdot ij}$ l'erreur type de $\frac{r_{ih}\,r_{jh}}{r_{ij}}$, cette estimation est donnée par

$$r^2_{hG} = \left[\sum_{i,j} \frac{r_{ih}\,r_{jh}}{r_{ij}\,\sigma^2_{h\cdot ij}}\right] \div \sum_{i,j} \frac{1}{\sigma^2_{h\cdot ij}} \qquad (h \neq i \neq j;\ i,j = 1, \dots, k)$$

En supposant que les k caractères sont distribués selon une loi de Laplace-Gauss à k variables, nous avons montré [7] que l'estimation de ρ_{hG} ayant l'écart type minimum, était donnée par

$$(8) \quad r^2_{hG} = \left[\sum_{i,j} \frac{r_{ij}\, r_{jh}}{r_{ij}\, W_{h\cdot ij}} \right] \div \sum_{i,j} \frac{1}{W_{h\cdot ij}} \quad (h \neq i \neq j;\ i, j = 1, \ldots, k)$$

où

$$(9) \quad \sigma^2_{h\cdot ij} = \frac{\rho_{ih}\, \rho_{jh}}{\rho_{ij}} \left[(v^2_{ih} + v^2_{jh} + v^2_{ih}\, v^2_{jh})(1 + 3v^2_{ij}) + 5v^4_{ij} + \cdots \right]$$

$$= \frac{\rho_{ih}\, \rho_{jh}}{\rho_{ij}}\, W_{h\cdot ij}$$

$$v_{ih} = \frac{1 - r^2_{ih}}{\sqrt{n-1}\,.\,r_{ih}} \quad \text{coefficient de variation de } r_{ih}.$$

Si les caractères ne sont liés que par un facteur général, les sommes $\sum_{i,j}$ de la formule (8) sont étendues à tous les $\frac{1}{2}(k-1)(k-2)$ couples de caractères ij ne comprenant pas h.

S'il existe des facteurs de groupes, les sommes $\sum_{i,j}$ ne devront être étendues qu'aux couples ij tels qu'aucun des ρ_{ih}, ρ_{jh}, ρ_{ij} ne fasse intervenir de facteur de groupe.

Remarquons que les valeurs de W varient très fortement et donnent parfois à certains termes de la moyenne pondérée (8) des poids 1,000,000 de fois plus faibles qu'à d'autres termes.

6. LOI DE DISTRIBUTION DE L'ESTIMATION DU COEFFICIENT DE CORRÉLATION ENTRE UN CARACTERE ET LE FACTEUR GÉNÉRAL

L'estimation r_{hG} donnée par la formule (8) est obtenue au moyen d'une moyenne arithmétique pondérée, donc la loi de distribution de r_{hG} autour de ρ_{hG} tend vers une loi de Laplace-Gauss lorsque le nombre des termes de $\sum_{i,j}$ est grand.

L'erreur type de cette distribution est donnée par:

$$(10) \quad \sigma_{r_{hG}} = \left(\sum_{i,j} \frac{1}{\sigma^2_{h\cdot ij}} \right)^{-\frac{1}{2}} = \rho_{hG} \div \left(\sum_{i,j} \frac{1}{W_{h\cdot ij}} \right)^{\frac{1}{2}}$$

les sommes $\sum_{i,j}$ étant étendues aux même couples de caractères que les $\sum_{i,j}$ de la formule (8) qui a servi à faire l'estimation r_{hG}.

Notons que l'estimation r_{hG} du coefficient de corrélation ρ_{hG} étant distribuée selon une loi de Laplace-Gauss autour de ρ_{hG}, on pourra obtenir parfois des valeurs $|r_{hG}| > 1$.

7. ESTIMATION ET LOI DE DISTRIBUTION DU COEFFICIENT DE CORRÉLATION ENTRE UN CARACTERE ET CHACUN DES FACTEURS DE GROUPES

La formule (8) permet d'obtenir des estimations $r_{hG} = a_h$ des coefficients $\rho_{hG} = \alpha_h$ du facteur général dans le système (1). Par la relation (2) on obtiendra les premières corrélations résiduelles

$$\rho'_{hi} = \rho_{hi} - \alpha_h\alpha_i = \beta_h\,\beta_i + \gamma_h\gamma_i + \cdots$$

estimées par $r'_{hi} = r_{hi} - a_h a_i$: Les couples hi de caractères n'admettant pas de facteur de groupe entre eux doivent donner les premiers résidus r'_{hi} statistiquement nuls. Pour les couples de caractères ayant 1 ou plusieurs facteurs de groupes entre eux, on analysera les résidus r'_{hi} de la même manière qu'on avait analysé précédemment les coefficients de corrélation r_{hi} observés. Pour cela on considérera le premier facteur de groupe comme le facteur général des r'_{hi} et les autres facteurs de groupes comme on avait précédemment considéré l'ensemble des facteurs de groupes. Les méthodes d'estimation des coefficients seront les mêmes, mais on remplacera dans la formule (9) W par W' et v_{jh} par $v'_{ih} = \sigma_{r'_{ih}}/r'_{ih}$ où

$\sigma^2_{r'_{ih}} = \sigma^2_{r_{ih}} + \sigma^2_{r_{hG}}\,r^2_{iG} + \sigma^2_{r_{iG}}\,r^2_{hG} + \sigma^2_{r_{iG}}\,\sigma^2_{r_{hG}}$[5] avec $\sigma_r = (1-r^2)/\sqrt{n-1}$

Les formules (8) et (10) deviendront respectivement pour le facteur de groupe B

$$(11) \qquad r^2_{hB} = \left[\sum_{i,j} \frac{r'_{ih}\,r'_{jh}}{r'_{ij}\,W'_{h\cdot ij}}\right] \div \sum_{i,j} \frac{1}{W'_{h\cdot ij}} \qquad (h \neq i \neq j)$$

$$(12) \qquad \sigma_{r_{hB}} = \left(\sum_{i,j} \frac{1}{\sigma'^2_{h\cdot ij}}\right)^{-\frac{1}{2}} = \rho_{hB} \div \left(\sum_{i,j} \frac{1}{W'_{h\cdot ij}}\right)^{\frac{1}{2}}$$

les sommes $\sum_{i,j}$ sont étendues à tous les couples ij introduisant le premier facteur de groupes suivants.

La loi de distribution de r_{hB} tend vers une loi de Laplace-Gauss pour les mêmes raisons que celle de r_{hG}.

---

[5] La valeur de $\sigma_{r'_{ih}}$ donnée par cette formule est un peu trop forte car les erreurs d'échantillonnage de r_{ih}, r_{iG} et r_{hG} ne sont pas indépendantes, toutes ces valeurs ayant été déterminées au moyen du même échantillon.

8. VÉRIFICATION DE L'EFFICACITÉ DE LA MÉTHODE

Une des principales qualités que doit présenter une méthode d'analyse factorielle est que, si un ensemble de caractères est réellement lié par une combinaison linéaire de facteurs indépendants, cette méthode doit permettre de retrouver le schéma de liaison et chacun des facteurs, et non pas d'autres schémas et d'autres facteurs, même si les mesures des caractères sont entachées d'erreurs.

D. Weinberg [13 et 14] Directeur-Adjoint du Laboratoire de Biométrie Humaine désirait qu'un tel contrôle des méthodes d'analyse factorielle fût fait et a organisé un schéma de 12 caractères 1, 2,..., 12 provenant d'un facteur général G, de 3 facteurs de groupes B, C, D, et de 12 facteurs spécifiques $S_1, \ldots, S_{12}$. Ensuite on a constitué artificiellement les résultats des mesures de 300 individus en déterminant pour chaque individu les 16 facteurs par le jet de dés à jouer. Ces résultats ont été composés selon le schéma pour former les individus fictifs qui ont servi à calculer les 66 coefficients de corrélation.

Ces coefficient de corrélation m'ont été remis par D. Weinberg sans aucune indication sur le nombre des facteurs ou sur la forme du schéma. J'ai utilisé la méthode ci-dessus pour analyser les coefficients de corrélation. L'étude des intervalles communs m'a permis de retrouver exactement le schéma à un facteur général et 3 facteurs de groupes qui avait été utilisé. Une première approximation [6] des coefficients de corrélation entre caractères et facteurs nous a donné des résultats voisins des coefficients théoriques de D. Weinberg (nous n'avions pas encore trouvé à ce moment l'estimation r_{hG} à écart type minimum). Voici les coefficients théoriques de corrélation entre caractères et facteurs, leurs estimations et les erreurs types de ces estimations calculées par les formules (8), (10), (11), et (12).

facteurs	Coefficients théoriques				Coefficients estimés				erreurs types			
	G	B	C	D	G	B	C	D	G	B	C	D
1	.577	.577	—	—	.572	.593	—	—	.034	.062	—	—
2	.236	.943	—	—	.271	.918	—	—	.034	.204	—	—
3	.943	.236	—	—	.939	.230	—	—	.016	.053	—	—
4	.436	.873	—	—	.480	.858	—	—	.032	.092	—	—
5	.577	—	.577	—	.500	—	.607	—	.035	—	.118	—
6	.236	—	.943	—	.205	—	.954	—	.050	—	.192	—
7	.918	—	.229	.229	.906	—	.243	.233	.024	—	.110	.105
8	.707	—	.471	.471	.660	—	.526	.554	.045	—	.106	.183
9	.943	—	—	.236	.941	—	—	.220	.020	—	—	.096
10	.236	—	—	.943	.223	—	—	.778	.050	—	—	.378
11	.949	—	—	—	.951	—	—	—	.015	—	—	—
12	.316	—	—	—	.365	—	—	—	.020	—	—	—

Les estimations des coefficients de corrélation entre caractères et facteurs étant distribuées autour des coefficients théoriques selon des lois de Laplace-Gauss, on peut tester pour le facteur général l'accord statistique entre valeurs théoriques et estimations par

$$\chi^2 = \sum_{i=1}^{12} [(r_{iG}-\rho_{iG})^2/\sigma^2_{r_{iG}} = 15.3$$

avec 12 degrés de liberté, ce qui correspond à une probabilité de 0.23 d'obtenir un accord moins bon. Ce resultat peut être considéré comme excellent.

Diverses applications de cette méthode à des données expérimentales ont été publiées [8].

CONCLUSION

L'introduction de la Statistique Mathématique en Analyse Factori· elle nous a permis de donner à celle-ci une forme tenant compte à tout moment de ce qu'elle analyse des données expérimentales, obtenues avec des erreurs de mesures sur des individus, et non pas des données mathématiques. La construction des graphiques des intervalles communs nous a donné une forme matériellement beaucoup plus simple que les tétrades pour repérer l'existence et localiser la position des facteurs de groupes. Nous avons enfin obtenu une estimation beaucoup plus précise du coefficient de corrélation entre caractère et facteur. L'usage de cette estimation à écart type minimum, nous a permis de connaître la distribution et l'erreur type de cette estimation; ce qui permettra à ceux qui utilisent l'Analyse Factorielle de connaître les variations de résultats qu'ils peuvent trouver d'un échantillon à un autre prélevé dans une même population, et donc d'interpréter avec plus de sécurité les résultats donnés par le calcul. Nous pensons que perfectionner ainsi l'outil qu'est l'analyse factorielle permettra son emploi dans un grand nombre de problèmes de biométrie, de météorologie, d'hydrologie, ou d'économétrie, de même que dans certaines fabrications industrielles, etc. où l'analyse de l'ensemble des caractères mesures semble actuellement ne pouvoir se faire utilement que par une décomposition en facteurs indépendants.

BIBLIOGRAPHIE

[1] G. Darmois, "Sur l'indétermination du facteur général dans la théorie de Spearman," *Mathematica*, XII (1936), p. 211-216.

[2] G. Darmois, *Les Mathématiques de la Psychologie*, Mémorial des Sciences Mathématiques, fasc. XCVIII, Paris: Gauthier-Villars, 1940, pp. 51.

[3] F. N. David, *Tables of the correlation coefficient*, Cambridge: The University Press, 1938, pp. XXXIX+55.

[4] P. Delaporte, "Une méthode d'analyse des corrélations," *Comptes Rendus Acad. Sci.*, Paris, t. 208 (1939), p. 1960.

[5] P. Delaporte, "Une méthode d'analyse des corrélations et son application," *Comptes Rendus Acad. Sci.*, Paris, t. 209 (1939), p. 142.

[6] P. Delaporte, "Vérification de l'efficacité d'une méthode d'analyse factorielle," *Comptes Rendus Acad. Sci.*, t. 220 (1945), p. 212.

[7] P. Delaporte, "Sur l'estimation des corrélations des caractères avec le facteur général et les facteurs de groupe et sur l'écart-type de cette estimation, en analyse factorielle," *Comptes Rendus Acad. Sci.*, t. 222 (1946), p. 525.

[8] P. Delaporte, "Prolongement de la méthode d'analyse factorielle de Spearman en utilisant la statistique mathématique," *Biotypologie*, IX (1946), No. 1-2.

[9] J. C. M. Garnett, "On certain independant factors in mental measurements, *Proc. Roy. Soc. London*, A, XCVI (1919), p. 91-111.

[10] K. Pearson and M. Moul, "The Mathematics of Intelligence. The Sampling Errors in the theory of a Generalized Factor," *Biometrika*, XIX (1927), p. 246-292.

[11] C. Spearman, "General Intelligence, Objectively Determined and Measured," *Amer. J. Psychol.*, XV (1904) p. 201-293.

[12] C. Spearman, *The Abilities of Man*, London: Macmillan, 1932.

[13] D. Weinberg, "Une expérience de contrôle des méthodes d'analyse factorielle," *Comptes Rendus Acad. Sci.*, Paris, t. 220 (1945) p. 214.

[14] D. Weinberg, "Contrôle expérimental des méthodes d'analyse factorielle," *Biotypologie*, VIII, No. 1-2, p. 56-74.

Résumé

The object of this paper is to show how by introduction of mathematical statistical methods it is possible to obtain a very large extension of factor analysis, first introduced by C. Spearman. It is assumed that the measurements of each variable can be explained by a linear function of a general factor G, of group factors, if necessary, and of the specific factor of the variable, and further that all these factors are statistically independant. By studying the sampling distribution of ratios of observed correlation coefficients, a confidence interval is determined for each ratio and for a given level of probability. This interval is obtained without computation on the chart of paragraph 4-2. The graphical interpretation of confidence intervals gives common interval graphics (par. 4-3) which are used instead of the classical tetrad differences test. These common interval graphics show which correlations

can be explained by a general factor only and also the correlations for which the introduction of group factors is necessary.

The next step is the estimation of the correlation coefficient between each variable and the general factor, obtained by a weighted arithmetical average with minimum variance of the sampling error (formula 8). The standard error (formula 10) and the sampling distribution of the estimate are calculated. This distribution is shown to be a normal law when the number of characters is large enough. These methods are generalized in order to obtain the corresponding results for group factors and to estimate the correlation coefficient between a character and each of the group factors.

Finally an experimental verification is given in order to show the efficiency of the author's method of factor analysis. From the observed correlation coefficients among 12 variables it has been possible to find out exactly, without any information about the scheme used in getting the data, that the true correlation had been caused by one general factor and 3 group factors. Correlation coefficients between variables and factors have been estimated and the χ^2 test has been applied, the whole group of errors of estimation has been shown to be in agreement with the computed standard errors.

THE MEANS OF SAMPLES

by Corrado Gini

University of Rome (Italy)

While much has been written on the influence of the limited number of observations on the value of indices of variability and relation, its influence on means, so far as I know, has never been studied. This probably depends on the fact that the expected value of the arithmetic mean—which is the one usually considered—of the sample, evidently coincides with the value of the mean of the collective phenomenon from which the sample is drawn.

But what of the other means? The question may have been discussed without my knowledge. Indeed, statistical literature has now taken such vast proportions that no one can presume to be acquainted with it all. So large is it indeed, that when the problem is a simple one, as in the present case, it takes less time to solve it than to find out whether it has been solved elsewhere.

I

Let us define first of all what we mean by "sample." A sample is a group of elements selected at random from the mass or collection of elements that compose the collective or mass phenomenon. Some call this mass "population," some "parent population," and some "universe." The two first expressions call to mind demographic phenomena or phenomena allied thereto. The last calls to mind the idea of the infinite and is therefore advisable only when the number of elements entering into the collective phenomenon may be supposed, by abstraction, to be infinite. But when the sample is taken from a positive mass of elements, necessarily finite in number, the word "mass" seems to me preferable, and I shall adhere to it.

The random selection of the sample may be made in several ways. Each element selected may be returned to the mass before the selection of another, or it may be kept separate from it. In the first case, we speak of *random selection with replacement*; and in the latter of *random selection without replacement*, meaning thereby "replacement of the selected element." But the entire sample whose elements are selected without replacement may, when it has been completed, either be or not be returned to the mass before another sample is drawn. We shall therefore speak of "replacement (or non-replacement) of the sample."

When we speak of "a sample selected at random," reference is generally made to a selection made by chance without replacement of the elements, but with replacement of the sample, and in this note we also will use the expression in that sense. In any case, in the problem we are discussing, the conclusions would not vary, as may be readily understood, even supposing the selection were made with replacement of the elements, or with non-replacement of sample as well as elements.

II

I shall limit my study to the firm basal monoplanar means, *i.e.*, to those in which the elements from which the mean is drawn are all present, and all as bases, and only in the numerator. The most comprehensive formula so far proposed for firm basal monoplanar means is that of the *combinatorial power mean*:

$$(1) \qquad {}^bM^p = \sqrt[bp]{\frac{1}{\binom{m}{b}} \sum_{l=1}^{\binom{m}{b}} P_l^b \, a_i^p}$$

in which m is the number of the elements of the mass, a_i $(i = 1, 2, \ldots, m)$ a generic element of the mass; p any real exponent to which the m elements of the mass are raised; $\binom{m}{b}$ the number of the combinations of the m elements b by b, while Σ as usual indicates the sum of the terms and P_l^b the product of the b elements in the combination.

If $b = m$, the combinatorial power mean is reduced to the geometric mean ; if $b = 1$, it is reduced to the *power mean* given by the well-known formula:

$$(2) \qquad M^p = \sqrt[p]{\frac{1}{m} \sum_{i=1}^{m} a_i^p}$$

If $p = 1$, the combinatorial power mean is reduced to the formula, also well known, of the *combinatorial mean*:[1]

$$(3) \qquad {}^bM = \sqrt[r]{\frac{1}{m} \sum_{l=1}^{\binom{m}{b}} P_l^b \, a_i}$$

Two combinatorial power means are described as *homograde* when the product bp in them is equal, and are described as *bihomograde* when both b and p are equal. All the bihomograde means are therefore homograde, but there are some homograde means that are not bihomograde. Thus two biquadratic means (in which $b = 1$, $p = 4$) are bihomograde (and therefore also homograde), as two combinatorial power means in which the terms raised to the second power are taken 4 by 4 (therefore $b = 4$, $p = 2$) are bihomograde. On the other hand, a biquadratic mean and a combinatorial mean whose terms are taken 4 by 4, and again a combinatorial mean whose terms raised to the square are taken 2 by 2, are homograde (in all three cases $bp = 4$) but not bihomograde.

---

[1] See, for all this, the article, "Di una formula comprensiva delle medie," *Metron*, vol. 13 (1938).

To determine the relations between the expected value of a certain mean of the sample and the value of the homograde or bihomograde mean of the mass, I will state four lemmas of which some have already been demonstrated while others can easily be demonstrated.

LEMMA 1: *The value of a power mean* M^{p} *for positive values of the terms* a_{i} *increases with the exponent* p.[2]

The geometric mean is, as is known, intermediate between the arithmetic and harmonic, and this is in keeping with the fact that, for p tending to zero, the power mean tends toward the geometric mean. Therefore the geometric mean may be indicated by the symbol M^0.

The following inequalities can therefore be established:

$$(4) \qquad \ldots M^{-2} < M^{-1} < M^0 < M^1 < M^2 < M^3 < M^4 \ldots$$

which show us that the biquadratic mean is larger than the cubic which is larger than the quadratic, and this than the arithmetic, and this than the geometric, and this than the harmonic, and this than the biharmonic.

LEMMA 2: *A power mean of the mass is equal to the homograde mean of the means (all being homograde as regards one another) of all the possible samples which may be drawn from the mass.*

One may draw from a mass of m elements $\binom{m}{s}$ different samples, each with s elements. If we indicate by ${}_mM^p$ the power mean of degree p of the mass, by ${}_mM^p_s$ the power mean of degree p of a sample of the said mass (in which $s = 1, 2, 3, \ldots, m-1$), and by $\mathfrak{M}^p \cdot ({}_mM^p_s)$ the power mean of degree p of the $\binom{m}{s}$ means ${}_mM^p_s$, the lemma tells us that[3]

$$(5) \qquad \mathfrak{M}^p({}_mM^p_s) = {}_mM^p$$

---

[2] This has been demonstrated by O. Dunkel, "Generalized Geometric Means and Algebraic Equations," *Annals of Mathematics*, 1909.

[3] Professor G. Usai has demonstrated this for the special cases of arithmetic, harmonic, and quadratic means. See G. Usai, "Alcune considerazioni sulle medie," *Bollettino dell'Accademia Gioenia*, Catania, 1940; "Proprietà combinatoria di certe medie—," *Esercitazioni matematiche*, publications of the "Circolo matematico di Catania," Vol. 12 (1940). For another demonstration, see G. Pompilj, "Sulla

Let us express this relation by saying that the power mean is *transitive as related to the sample.*[4]

Lemma 3: *The geometric mean of the mass is equal to the geometric mean of the geometric means of all the possible samples that can be drawn from the mass.*

That is to say[5]

$$\mathfrak{M}^0({}_mM_s^0) = {}_mM^0 \tag{6}$$

This property was to be expected because the geometric mean is the limit of the power mean in the special case where $p = 0$.

Therefore the *geometric mean is also transitive as related to the sample.*

Lemma 4: *A combinatorial power mean of the mass is equal to the homograde power mean of the combinatorial power means (which are bihomograde as regards one another) of all the possible samples which may be drawn from the mass.*

Let ${}_m^bM^p$ be the combinatorial power mean of degree bp of the m elements of the mass, raised each to the p power, and combined b by b; ${}_m^bM_s^p$ the combinatorial power mean of degree bp of a sample (in which $s = 1, 2, \ldots, m$) and $\mathfrak{M}^{bp}({}_m^bM_s^p)$ the power mean of degree bp of the $\binom{m}{s}$ means of the samples. The lemma says[6]

$$\mathfrak{M}^{bp}({}_m^bM_s^p) = {}_m^bM^p \tag{7}$$

---

media geometrica o sopra un indice di mutabilità calcolato mediante un campione —, "*Memorie della Societa Italiana delle Scienze, detta dei XL,* Series III, Vol. 26. It was this communication submitted by Dr. Pompilj to the Seminary of the Institute of Statistics of the University of Rome (Faculty of Statistical, Demographic, and Actuarial Sciences) which suggested to me the present paper.

The general demonstration is easy. From the definition of a sample it follows that each element of the mass has the same probability of being an element of the sample so that in the body of the possible $\binom{m}{s}$samples each element of the mass will enter $\binom{m}{s}\frac{s}{m}$ times, *i.e.*, $\binom{m-1}{s-1}$ times: now, if of the $\binom{m}{s}$ power means of degree p of all the possible samples we take the power mean of degree p, we shall obtain the power mean of degree p of the m elements of the mass, each of which is repeated $\binom{m-1}{s-1}$ times. Evidently, the value of this mean is the same as that of the power mean of degree p of the m elements, if each be taken once only, *i.e.*, of the power mean of degree p of the mass.

[4] It is commonly said that a mean is *transitive* when the mean of s means, each of n elements, is equal to the mean of the sn elements.

[5] G. Usai, "Proprietà combinatoria di certe medie," *loc. cit.*, demonstrates this.

[6] So far as I know, the demonstration has never been given, It is analogous to that given under note 3. Indeed, if of the $\binom{m}{s}$ combinatorial power mean of degree bp relating to all the possible samples the power mean of degree bp be made,

It should be noted that the lemma says "it is equal to the homograde power mean," and not "it is equal to the bihomograde combinatorial power mean." Therefore, the *combinatorial power means are not transitive as related to the sample.*

Let us remember that the expected value of a certain mean of the samples which we will indicate by $E({}_{m}M_{s}^{p})$ for the power means, and $E({}_{m}^{b}M_{s}^{p})$ for the combinatorial power means, is nothing but the arithmetic mean of the means, homograde or respectively bihomograde as regards one another, of all the possible samples which may be drawn from the mass. We might therefore write

$$E({}_{m}M_{s}^{p}) = \mathfrak{M}^{1}({}_{m}M_{s}^{p}) \tag{8}$$

$$E({}_{m}^{b}M_{s}^{p}) = \mathfrak{M}^{1}({}_{m}^{b}M_{s}^{p}) \tag{9}$$

Now, from Lemma 1 we obtain

$$\mathfrak{M}^{-2}({}_{m}M_{s}^{p}) < \mathfrak{M}^{-1}({}_{m}M_{s}^{p}) < \mathfrak{M}^{0}({}_{m}M_{s}^{p}) < \mathfrak{M}^{1}({}_{m}M_{s}^{p}) < \mathfrak{M}^{2}({}_{m}M_{s}^{p}) < \\ < \mathfrak{M}^{3}({}_{m}M_{s}^{p}) < \mathfrak{M}^{4}({}_{m}M_{s}^{p})$$

$$\mathfrak{M}^{-2}({}_{m}^{b}M_{s}^{p}) < \mathfrak{M}^{-1}({}_{m}^{b}M_{s}^{p}) < \mathfrak{M}^{0}({}_{m}^{b}M_{s}^{p}) < \mathfrak{M}^{1}({}_{m}^{b}M_{s}^{p}) < \mathfrak{M}^{2}({}_{m}^{b}M_{s}^{p}) < \\ < \mathfrak{M}^{3}({}_{m}^{b}M_{s}^{p}) < \mathfrak{M}^{4}({}_{m}^{b}M_{s}^{p})$$

and therefore, from (5), (6), and (8),

$$E({}_{m}M_{s}^{p}) = {}_{m}M^{p}$$

for the arithmetic mean, *i.e.*, for $p = 1$;

$$E({}_{m}M_{s}^{p}) > {}_{m}M^{p} \tag{10}$$

for the geometric and harmonic means and in general for all those for which $p < 1$;

$$E({}_{m}M_{s}^{p}) < {}_{m}M^{p} \tag{11}$$

we shall obtain a new combinatorial power mean of degree bp which includes $\binom{m}{s}\binom{s}{b}$ combinations, each formed of b elements and in each of which every element of the mass has the probability of appearing $\frac{s}{m} \cdot \frac{b}{s} = \frac{b}{m}$. This combinatorial power mean has indeed the same value as the combinatorial power mean of degree bp of the mass which includes $\binom{m}{b}$ combinations each formed of b elements and in each of which every element of the mass has the probability $\frac{b}{m}$ of appearing. Indeed the two means only differ in their composition owing to the fact that each of the combinations that appears in the second is repeated in the first;
$\binom{m}{s}\binom{s}{b} : \binom{m}{b} = \binom{m-b}{m-s}$ times.

for the quadratic, cubic, and biquadratic means and, generally speaking, for all those for which $p > 1$; and from (7) and (9),

$$E({}_m^b M_s^p) = {}_m^b M^p \qquad \text{for } p = 1/b;$$

(12) $$E({}_m^b M_s^p) > {}_m^b M^p \qquad \text{for } p < 1/b; \text{ and}$$

(13) $$E({}_m^p M_s^p) < {}_m^b M^p \qquad \text{for } p > 1/b \text{ and}$$

more especially for $p = 1$, in which last case we have the combinatorial power mean.

As is seen, *it is only in special cases, i.e., only for the power means or for the combinatorial power means of degree one, that the expected value of the sample mean is equal to the expected value of the homograde mean of the mass.* In the other cases it is larger for some means and smaller for others. To be precise, the expected value of the sample mean is larger for all the power means or combinatorial power means of degree lower than one, and smaller for all those of degree higher than one.

As the value of the power means or combinatorial power means increases with the value of the exponent p, it follows that with the variation of the exponent p, the power mean or combinatorial power mean of the sample varies less than that of the mass. This conclusion helps to solve a practical point that was called to my attention, in relation to the compilation of price index numbers.

Many writers, after having calculated the price index numbers for individual commodities prefer, in order to obtain therefrom the average index number of prices of said commodities, to calculate the geometric rather than the arithmetic mean of the index numbers for the individual commodities. Now, suppose that the geometric means have thus been calculated for the individual commodities entering into the various groups of goods (as for instance, agricultural products, ores, semi-manufactures, and finished products). Should the geometric mean or the arithmetic mean of these group index numbers be calculated in order to obtain a general price index number of the commodities belonging to all the groups? The reply is: the geometric mean, inasmuch as it is the geometric mean of the group index numbers which gives the geometric mean of the index numbers of all the individual commodities belonging to all the groups.

III

The inequalities (10), (11), (12), and (13) are more marked the smaller the sample taken from the mass. This is easily understood when it is remembered that given a mass m and two samples s' and s'', for which $s' < s''$, s'' may be considered as a mass in relation to s' so that, between the means of s' and s'' the difference occurs in the same direction as between the means of s'' and m, and consequently the difference between the means of s' and m is equal in direction and greater in intensity than the difference between the means of s'' and m.

Similarly, one can understand that with the variation of the exponent p, the power mean or combinatorial power mean varies less, the smaller the sample.

IV

The combinatorial power mean ${}^b_m M^p$ for $b = m$ becomes equal to the geometric mean ${}_m M^0$, whatever p may be; for $b = 1$ it becomes equal to the power mean ${}_m M^p$ whose value varies with the variations in the value of p. Therefore, in conformity with the principle of continuity, it is to be expected that the value ${}^b_m M^p$, m being constant, will vary all the more with the variation of p, the smaller is b.

V

The combinatorial power mean of degree bp of a sample may be reduced to a homograde power mean of the geometric means of the $\binom{m}{b}$ samples, each formed of b elements, that may be drawn from the mass.

It is indeed

$$ {}^b_m M^p = \mathfrak{M}^{bp}({}_m M^0_b) \tag{14} $$

Now the second term of the equation may be generalized by taking, instead of the geometric mean M^0, and the power mean $\mathfrak{M}^{bp}$, any two other means. The same result is obtained by generalizing (9) by replacing the arithmetic mean $\mathfrak{M}^1$ by a combinatorial power mean. The expression thus obtained

$$ {}^e\mathfrak{M}^d({}^b_m M^p_s) \tag{15} $$

expresses a true mean and supplies an expression of the firm basal monoplanar mean which is evidently more general than (1).

VI

Let us check, by an example, the conclusions reached above. Take a mass of five terms: 1, 4, 5, 10, 80.

Table I illustrates the differences between the power mean of the mass, ${}_m M^p$, and the expected value of the homograde power mean of the sample $E({}_m M^p_s)$ for $s = 1, 2, 3, 4$. As will be seen by running over the several columns, the arithmetic mean (by $p = 1$) of the mass ($= 20$), is equal to the expected value of the arithmetic mean of the sample, whatever may be the number (1, 2, 3, 4) of elements it includes.

For the quadratic, cubic, biquadratic means ($p = 2, 3, 4$) the expected value of the mean of the sample is found instead to be smaller than the homograde mean of the mass, and the inequality is all the greater the smaller the sample (for instance, for $p = 3$, we have $29.006 < 35.487 < 41.648 < 46.820$). On the other hand, for the geometric ($p = 0$), harmonic ($p = -1$), and biharmonic ($p = -2$) means, the expected value of the mean of the sample is found to be greater than the value of the homograde mean of the mass and the inequality is all the greater the smaller the sample (for instance, for $p = -2$, we have $4.831 > 3.484 > 2.729 > 2.12$).

Table I also shows that, with the variation in the value of the exponent p, the value of the power mean varies more in the mass than in the sample, and the smaller the sample the less it varies. Thus, when p passes from -2 to 4, the power mean varies in the mass from 2.1 to 53.5; while in a sample of 4 elements its expected value varies from 2.7 to 46.7, in one of 3 elements from 3.5 to 39.2, in one of two elements from 4.8 to 30.5, and finally it coincides always with the arithmetic mean in the extreme case in which the sample consists of one element only.

Tables Ia, Ib, and Ic, show in detail the calculation of the expected value of the power mean of the sample $E({}_mM_s^p)$ in the respective cases $s = 4$, $s = 3$, $s = 2$.

Tables II and III illustrate the differences between the combinatorial power mean of the mass ${}^b_mM^p$ and the expected value of the bihomograde means of the sample $E({}^b_mM^p_s)$ for $m = 5$, $b = 2$, and $s = 4, 3,$ and 2 (Table II); and for $m = 5$, $b = 3$, and $s = 4$ and 3 (Table III). Both in Table II and in Table III the expected value of the mean of the sample is greater than the value of the mean of the mass for the geometric, harmonic, and biharmonic means, and smaller for all the others, including the arithmetic mean, as $1/b = \frac{1}{2}$ in Table II and $1/b = \frac{1}{3}$ in Table III, and therefore in both cases $0 < 1/b < 1$.

It should also be noted that the inequalities increase as the number of elements in the sample decreases. Thus, for instance, in Table II, for $p = 2$ and $b = 2$, we have $10.028 < 13.268 < 15.784 < 17.385$, while for $p = -2$ and $b = 2$, we have $10.028 > 4.396 > 3.598 > 3.132$.

Tables II and III also illustrate how, with the variation of the exponent p the value of the combinatorial power mean varies more in the mass than in the sample and in the sample the less the smaller it is, until the said value of the combinatorial power mean becomes constant and equal to the geometric mean when the sample is equal to the number of elements entering into the combination. Thus for $b = 3$ (Table III) the value of ${}^3_5M^p$ passes, with the variation of p from -2 to 4, from

3.8 to 13.5, while the value of $E({}_5^3M_s^p)$ for $s = 4$ passes from 4.5 to 11.5 and for $s = 3$ remains always equal to 8.1.

Tables IIa, IIb, and IIIa show the details of the calculation of the expected value of the combinatorial power mean of the sample, respectively for $b = 2$ and $s = 3$, for $b = 2$ and $s = 4$, and for $p = 3$ and $s = 4$.

Table IV illustrates how the values that the combinatorial power mean of the mass ${}_m^bM^p$ acquires with the variations of p vary all the less the more b approaches m and at last become equal for $b = m$. Thus, as p passes from -2 to 4, the value of ${}_m^bM^p$ passes from 2.1 to 53.5 for $b = 1$, from 3.1 to 21.4 for $b = 2$, from 3.8 to 13.5 for $b = 3$, from 4.6 to 10.2 for $b = 4$, and remains constant at 6.931 for $b = 5$.

Lastly Tables V and VI illustrate the similar relations between the expected value of the combinatorial power mean of the sample $E({}_m^bM_s^p)$ and the value of b in the case of samples of 4 and 3 elements.

TABLE I

Comparison between the values of ${}_mM^p$ and the respective values of $E({}_mM_s^p)$ for different values of p and s

p	-2	-1	0	1	2	3	4
${}_mM^p$	2.12	3.2	6 931	20	36.172	46.820	53.503
$E({}_mM_s^p)$ for:							
$s = 4$	2.729	3.682	7.361	20	33.457	41.648	46.701
$s = 3$	3.484	4.396	8.146	20	29 998	35.857	39.200
$s = 2$	4.831	5.869	10.028	20	26.057	29.006	30.455
$s = 1$	20	20	20	20	20	20	20

TABLE Ia: *Various values of ${}_mM_s^p$ in all the possible samples for $m = 5$, $s = 4$, and for different values of p, and their arithmetic mean $E({}_5M_4^p)$*

Elements of the sample	p						
	-2	-1	0	1	2	3	4
1, 4, 5, 10	1.896	2.581	3.761	5	5.959	6.676	7.222
1, 4, 5, 80	1.905	2.735	6.325	22.5	40.131	50.263	56.569
1, 4, 10, 80	1.931	2.936	7.521	23.75	40.363	50.432	56.572
1, 5, 10, 80	1.952	3.048	7.953	24	40.392	50.434	56.572
4, 5, 10, 80	5.959	7.111	11.247	24.75	40.438	50.436	56.572
$E({}_5M_4^p)$	2.729	3.682	7.361	20	33.457	41.648	46.701

TABLE Ib: *Various values of* ${}_mM_s^p$ *in all the possible samples for* $m = 5$, $s = 3$, *and their arithmetic mean* $E({}_5M_s^p)$

Elements of the sample	p						
	−2	−1	0	1	2	3	4
1, 4, 5	1.650	2 069	2.714	3.333	3.742	3.986	4.141
1, 4, 10	1.672	2.222	3.420	5	6.245	7.080	7.647
1, 4, 80	1 680	2.376	6.840	28.333	45.249	55.47	60.79
1, 5, 10	1.690	2.308	3.684	5.333	6.481	7.213	7.714
1, 5, 80	1.698	2.474	7.368	28.667	45.282	55.47	60.79
1, 10, 80	1.723	2.697	9.283	30.333	46.551	55.51	60.79
4, 5, 10	5.164	5.455	5.848	6.333	6.856	7.345	7.760
4, 5, 80	5.406	6.486	11.696	29.667	46.336	55.48	60.79
4, 10, 80	6.426	8.276	14.736	31.333	46.605	55.51	60.79
5, 10, 80	7.734	9.600	15.874	31.667	46.637	55.51	60.79
$E({}_5M_3^p)$	3.484	4.396	8.146	20	29.998	35.857	39.20

TABLE Ic: *Various values of* ${}_mM_s^p$ *in all the possible samples for* $m = 5$, $s = 2$, *and their arithmetic mean* $E({}_5M_2^p)$

Elements of the sample	p						
	−2	−1	0	1	2	3	4
1,4	1.372	1.600	2	2.5	2.915	3.191	3.367
1, 5	1.387	1.667	2.236	3	3.606	3.979	4.106
1, 10	1.407	1.818	3.162	5.5	7.106	7.940	8.409
1, 80	1.414	1.975	8.944	40.5	56.57	63.50	67.27
4, 5	4.417	4.444	4.472	4.5	4.528	4.555	4.581
4, 10	5.252	5.714	6.225	7	7.616	8.103	8.462
4, 80	5.650	7.619	17.889	42	56.639	63.50	67.27
3, 10	6.325	6.667	7.071	7.5	7.906	8.255	8.537
5, 80	7.057	9.412	20	42.5	56.679	63.50	67.27
10, 80	14.03	17.778	28.284	45	57.009	63.54	67.28
$E({}_5M_2^p)$	4.831	5.869	10.028	20	26.006	29.006	30.655

TABLE II

Comparison between the values of ${}_m^bM^p$ *and the respective values of*
$E({}_m^bM_s^p)$ *for* $m = 5$, $b = 2$, *for different values of* s *and* p

p	−2	−1	0	1	2	3	4
${}_5^2M^p$	3.132	3.88	6.931	13.149	17.385	19.838	21.436
$E({}_5^2M_s^p)$, for:							
$s = 4$	3.598	4.430	7.361	12.460	15.784	17.155	18.183
$s = 3$	4.396	5.488	8.146	11.547	13.268	14.199	14.749
$s = 2$	10.028	10.028	10.028	10.028	10.028	10.028	10.028

TABLE IIa: *Various values of* ${}_m^bM_s^p$ *in all the possible samples for* $m = 5$, $s = 4$, *and* $b = 2$, *and for different values of* p, *and their arithmetic mean* $E({}_5^2M_4^p)$

Elements of the sample	p						
	−2	−1	0	1	2	3	4
1, 4, 5, 10	2.323	3.050	3.761	4.637	5 274	5.664	5.914
1, 4, 5, 80	2.747	3.403	6.325	11.754	14.553	15.909	16.690
1, 4, 10, 80	3.008	3.913	7.521	14.457	18.806	21.204	22.681
1, 5, 10, 80	3.300	4.224	7.953	14.972	19.163	21.402	22.781
4, 5, 10, 80	6.611	7.559	11.247	16.482	19.718	21.598	22.849
$E({}_5^2M_4^p)$	3.598	4.430	7.361	12.460	15.503	17.155	18.163

TABLE IIb: *Various values of* ${}_m^bM_s^p$ *in all the possible samples for* $m = 5$, $s = 3$ *and* $b = 2$, *and for different values of* p, *and their arithmetic mean* $E({}_5^2M_3^p)$

Elements of the sample	p						
	−2	−1	0	1	2	3	4
1, 4, 5	2.301	2.449	2.714	3.109	3.482	3.638	3.901
1, 4, 10	2.521	2.828	3.420	4.234	4.890	5.281	5.516
1, 4, 80	2.620	3.361	6.840	11.60	13.8	14.933	15.601
1, 5, 10	2.778	3.062	3.684	4.761	5.439	5.897	6.165
1, 5, 80	2.940	3.735	7.368	12.71	15.035	16.676	17.44
1, 10, 80	4.145	5.136	9.283	17.22	21.545	23.556	24.65
4, 5, 10	5.400	5.619	5.848	6.055	6.223	6.351	6.448
4, 5, 80	5.417	7.344	11.696	15.70	17.204	17.841	18.20
4, 10, 80	6.577	10.11	14.736	19.66	22.316	23.797	24.73
5, 10, 80	9.261	11.24	15.874	20.41	22.742	24.019	24.84
$E({}_5^2M_3^p)$	4.396	5.488	8.146	11.547	13.268	14.199	14.749

TABLE III

Comparison between the values of ${}_m^bM^p$ *and the respective values of* $E({}_m^bM_s^p)$ *for* $m = 5$, $b = 3$, *for different values of* s *and* p

p	−2	−1	0	1	2	3	4
${}_5^3M^p$	3.844	4.455	6.931	9.994	119.92	12.933	13.504
$E({}_5^3M_s^p)$, for:							
$s = 4$	4.470	5.368	7.361	9.394	10.519	11.134	11.498
$s = 3$	8.146	8.146	8.146	8.146	8.146	8.146	8.146

TABLE IIIa: *Various values of ${}_m^bM_s^p$ in all the possible samples for $m = 5, s = 4$, and $b = 3$, and their arithmetic mean $E({}_5^3M_4^p)$*

Elements of the sample	*p*						
	−2	−1	0	1	2	3	4
1, 4, 5, 10	2.67	3.42	3.761	4.264	4.725	5.027	5.212
1, 4, 5, 80	3.416	4.143	6.325	8.363	9.436	10.052	10.425
1, 4, 10, 80	4.296	5.127	7.521	10.291	11.834	12.656	3.133
1, 5, 10, 80	4.627	5.503	7.953	10.949	12.702	13.622	14.144
4, 5, 10, 80	7.341	8.647	11.247	13.104	13.899	14.313	14.574
$E({}_5^3M_4^p)$	4.470	5.368	7.361	9.394	10.519	11.134	11.498

TABLE IV

Different values acquired by ${}_m^bM^p$ with the variations in the values of b and p

b	*p*						
	−2	−1	0	1	2	3	4
1	2.12	3.2	6.931	20	36.172	46.820	53.503
2	3.132	3.886	6.931	13.149	17.385	19.838	21.436
3	3.844	4.455	6.931	9.994	11.992	12.933	13.504
4	4.586	5.318	6.931	8.409	9.321	9.855	10.205
5	6.931	6.931	6.931	6.931	6.931	6.931	6.931

TABLE V

Different values acquired by $E({}_m^bM_s^p)$ with the variations in the values of b and p for $s = 4$

b	*p*						
	−2	−1	0	1	2	3	4
1	2.729	3.682	7.361	20	33.457	41.648	46.701
2	3.598	4.430	7.361	12.460	15.784	17.155	18.183
3	4.470	5.368	7.361	9.394	10.519	11.134	11.498
4	7.361	7.361	7.361	7.361	7.361	7.361	7.361

TABLE VI

Different values acquired by $E({}^b_m M^p_s)$ *with the variations in the values of b and p for* $s = 3$

b	p: −2	−1	0	1	2	3	4
1	3.484	4.396	8.146	20	29.998	35.857	39.200
2	4.396	5.488	8.146	11.547	13.268	14.199	14.749
3	8.146	8.146	8.146	8.146	8.146	8.146	8.146

Résumé

Dans cette note, on montre les relations qui interviennent entre les moyennes d'une masse (population ou univers) et l'espérance mathématique des moyennes correspondantes d'un échantillon.

A cet effet, nous rappelons les formules de la

1. *Moyenne de puissances*

$$M^p = \sqrt[p]{\frac{1}{m}\sum_{i=1}^{m} a_i^p} \tag{1}$$

formule dans laquelle m indique le nombre des termes, a_i $(i = 1, 2, \ldots, m)$ le terme général i^{eme} et p fournit le *degré* de la moyenne;

2. *Moyenne combinatorielle*

$$ {}^bM = \sqrt{\frac{1}{\binom{m}{b}} \sum_{l=1}^{\binom{m}{b}} P_l^b a_i} \tag{2}$$

formule dans laquelle $\binom{m}{b}$ indique le nombre de combinaisons des termes m, b à b, $P_l^b a_i$ le produit des b éléments compris dans la combinaison l^{eme} et b fournit le *degré* de la moyenne;

3. *Moyenne combinatorielle de puissances*

$$(3) \qquad {}^bM^p = \sqrt[bp]{\frac{1}{\binom{m}{b}} \sum_{l=1}^{\binom{m}{b}} P_l^b \, a_l^p}$$

formule dans laquelle bp fournit le *degré* de la moyenne.

(1) et (2) constituent des cas particuliers de (3) pour $b = 1$ et respectivement pour $p = 1$.

Si nous désignons par

$$_mM^p \qquad {}_m^bM \qquad {}_m^bM^p$$

les moyennes pour la masse comprenant m éléments et par

$$E({}_mM_s^p) \qquad E({}_m^bM_s) \qquad E({}_m^bM_s^p)$$

les espérances mathématiques des moyennes respectives pour un échantillon comprenant s éléments, on peut démontrer que

$$E({}_mM_s^p) \gtreqless {}_mM^p \qquad E({}_m^bM_s) \gtreqless {}_m^bM \qquad E({}_m^bM_s^p) \gtreqless {}_m^bM^p$$

selon que le degré (respectivement p, b, bp) de la moyenne est $\lesseqgtr 1$.

On peut déduire de ce qui précède, en particulier, que l'espérance mathématique de la moyenne géometrique et harmonique de l'échantillon est supérieure à la moyenne géometrique et harmonique respective de la masse, tandis que l'espérance mathématique de la moyenne quadratique, cubique, et biquadratique de l'échantillon est inférieure à la moyenne quadratique, cubique et biquadratique respective de la masse.

On peut démontrer aussi que:

1) la différence entre la moyenne de la masse et l'espérance mathématique de la moyenne respective de l'échantillon est d'autant plus grande que l'échantillon est petit;

2) avec la variation de l'exposant p l'espérance mathématique de la moyenne de puissances ou de la moyenne combinatorielle de puissances de l'échantillon varie moins que l'espérance mathématique de la moyenne respective de la masse, et plus petit l'échantillon, moins elle varie;

3) si le nombre de termes (m) est égal, la valeur de la moyenne combinatorielle de puissances ${}_m^bM^p$ varie avec la variation de p d'autant plus que b est petit.

ON RANDOM VARIATIONS IN STATISTICAL DATA*

by Wilhelm Winkler

University of Vienna (Austria)

Note: This study is based on the *Lexis method* of measuring the dispersion of relatives and other statistical measurements. That method, which appeared to belong to the established stock of statistical methods and was represented in every textbook on the theory of statistics or the theory of chance, has been attacked by some authors in more recent times as being unreliable and obsolete.[1]

Nevertheless, I believe that the Lexis method, though it may have its faults, is indispensable in its ease of understanding and handling, and can never be replaced by more refined methods which are too exacting to be of broad practical use.

The lack of literature prolonged for the third year after Austria's "liberation" may also be the explanation and excuse, if it is found that the methods proposed in this study have been found by others before me. They are simple and obvious enough and may well have occurred to other investigators. Under those conditions this study is only meant as an unpretending contribution to the problem of sampling and as a basis for discussion, and I am still more willing than ever to accept correction and instruction if needed.

THE PROBLEM

In spite of many excellent and ingenious investigations, sometimes requiring difficult mathematical work, there is still disagreement about the fundamental problem whether or not it is correct to apply the formulae of the theory of chance to statistical data; or, otherwise expressed, whether and to what extent experience reflected in statistical data follows the pattern given in the theory of chance.

The followers of the liberal line of thought answer in the affirmative: They can be recognized by their practice of setting $\pm$ E, that is the probable error of the theory of chance, next to all proportions they can study, while they argue that actual life with its working of a multiplicity of individual causes, besides the general ones, may rightly be assumed to be a complete parallel to the games of chance. The followers of the strict line of thought assert there is no justification for the assumption of that parallel except in those few cases of time series composed of proportions based on a large number of observations each year for which "normal dispersion" (according to Lexis) has been proved.

---

* This paper is an abridged version of that presented at the conferences. A full version in German has been published in *Statistische Vierteljahresschrift*, I, Vienna, 1948.

[1] See Oscar Anderson, *Einfuhrung in die mathematische Statistik*, Vienna, 1935, p. 300 *ff.*, and literature quoted there.

I believe that neither line of thought is quite correct and that there is a golden mean which better approaches the truth.

When examining the first line of thought, we must not forget the rigorous conditions the classical (and also the new) theories of chance set up for the validity of those formulae: The best possible mixture during the whole course of sample drawings, say from a bag filled with balls of different colors in given proportions; unvarying composition in the bag from sample to sample; no limitations upon the free working of chance; complete independence between successive drawings.

Changes in the basic probability from sample to sample increase the dispersion. They may be effected either by changing the composition of the colors of balls in our bag or by not fully mixing the contents so that draws may be alternately taken from different spheres of mixture in the bag. In samples taken from actual life changes have proved to be the rule, and stability the rare exception. In investigating time series of proportions we shall be unlikely to find normal dispersion, a fact which has already been stated by Lexis. But also the condition of perfect mixture seems rarely fulfilled in actual populations: Rural and urban conditions are greatly different; there are rich and poor quarters in towns; there are agglomerations of one sex in barracks, colleges, cloisters, etc., agglomerations of people exposed to a higher mortality in hospitals, of people with a greater tendency to emigration in sea-side places, etc. Thus it may happen that in taking successive samples out of the bag of social life we may, even with the general conditions remaining unchanged, get samples from different corners of our bag which is not well-mixed. Thus, we may have the impression of a change in the general conditions without its being actual.

Against these most frequent disturbing causes the last two, limitation of the free working of chance and dependence between successive drawings, seem to be of lesser importance with actual observations.

It follows from these considerations that the liberal line of thought seems to go too far and to need certain restrictions on its practice of uncritically applying probability formulae to actual series of proportions. On the other hand, the strict line of thought, recognizing the use of the probability formulae only in those cases where normal dispersion has been found, seems to me wrong too. The mistake seems to consist in denying the character of a probability to all those proportions which are not constant in time, as if that constancy in time were a constitutive quality of a probability. Nobody will deny that the probabilities of survival and of mortality given in a mortality table are specimens of probability, and no one will ignore the fact that those probabilities have undergone the greatest changes for the last seventy years in most countries. That mistake is an old one, and still older are its causes.

Therefore, we must not find an argument against the applicability of the well-known "limits of significance" σ or E or their multiples, in those changes of the basic probability from sample to sample. It was just this that many of Lexis' successors did, and do to this day.

The present question is how to eliminate the disturbing effect of time and bring given statistical figures more nearly in accord with the conditions of random sampling. The best approach would be to take samples at random of every statistical mass of data as long as the original order of data, *e.g.*, the order in which population lives together in lodgings, houses, villages, and towns, is preserved. To execute that idea we may arrange unsorted punch cards in divisions and subdivisions of shelves and determine by lot the numbers and sub-numbers from which cards are to be drawn at random in order to form samples from them. Possible limitations of chance or correlations between the units may be taken regard of, in order to imitate actual life as nearly as possible. Then the dispersion of these samples may be ascertained and compared with that of the Bernoullian distribution, using Lexis' coefficient of dispersion $Q = \sigma/\sigma_0$, in which σ means the (actual) standard deviation observed, σ_0 the (theoretical) standard deviation of the Bernoullian distribution. In this way we could obtain an idea of the effect of distorting causes of dispersion remaining after the exclusion of those of time, upon the dispersion of the proportions to be dealt with, *i.e.*, distortions caused by poor mixture of our material, or correlations, etc. Those σ's would give us the right measures for the probable errors of sampling appropriate to our material.

That way, though theoretically easy, seems to me hardly feasible, as every acting statistician will certify.

Therefore, I suggest another way of studying random variations in actual statistical figures: We may avail ourselves of figures given year by year, but not refer their variations to the average for the appropriate period of time, as Lexis did, but to the trend of their change in time, calculated by some method of graduation. This may be done either by the method of moving averages (by which terms at the beginning and ending of our series are lost), or by the method of least squares in connection with an assumption fitting the nature of that trend. Because of the progress of statistical studies on business cycles there is hardly a need for further justification of that device.

The first example refers to some of the figures used by Lexis himself in illustrating his method. Using these figures saves us one part of the calculation, as Q is already calculated in the Lexis way. Data are for the sex distribution of deaths for the ages 50 to 55 years in Belgium from 1841 to 1860. (See Table I.)

In the first column of Table I we find data on the sex distribution of deaths, not in the form in which Lexis used them—males to 1000 females—but transformed to probabilities. In the second column these figures are graduated in two segments from 1841 to the distinct culminating point in 1847 by an algebraic curve of the third degree, centering in the year 1844, $y = .471143 + .004846x + .005096x^2 + .001550x^3$; from there on by a straight line, centering in 1854, $y = \cdot542877 + \cdot002904x$. The graduation in two segments gave a better fit and was admissible, as here there was not the question of finding one expression for the whole period of time but of finding the best fit, if necessary in two or more segments. In column 3 we have the deviations of the observed from the graduated values, in column 5 from the general average of the observed values,

TABLE I

Proportion of Males among all Deaths of Ages 50 to 55 in Belgium[1]

Year	Proportion of males		$\lambda \times 10^4$ ($\lambda = y - y'$)	$\lambda^2 \times 10^8$	$\delta \times 10^4$ ($\delta = y - 0.5292$)	$\delta^2 \times 10^8$
	rough y	graduated y'				
	1	2	3	4	5	6
1841	0.4556	0.4576	− 20	400	−736	541,696
1842	0.4731	0.4674	+ 57	3249	−561	341,721
1843	0.4650	0.4688	− 38	1,444	−642	412,164
1844	0.4703	0.4711	− 8	64	−589	346,921
1845	0.4829	0.4836	− 7	49	−463	214,369
1846	0.5188	0.5156	+ 32	1,024	−104	10,816
1847	0.5750	0.5764	− 14	196	+458	209,764
1848	0.5185	0.5255	− 70	4.900	−107	11,449
1849	0.5323	0.5284	+ 39	1,521	+ 31	961
1850	0.5261	0.5313	− 52	2,704	− 31	961
1851	0.5396	0.5342	+ 54	2,916	+104	10,816
1852	0.5381	0.5371	+ 10	100	+ 89	7,921
1853	0.5467	0.5400	+ 67	4,489	+175	30,625
1854	0.5392	0.5429	− 37	1,369	+100	10,000
1855	0.5546	0.5458	+ 88	7,744	+254	64,516
1856	0.5560	0.5487	+ 73	5,329	+268	71,824
1857	0.5292	0.5516	−224	50,176	+ 0	0
1858	0.5532	0.5545	− 13	169	+240	57,600
1859	0.5633	0.5574	+ 59	3,481	+341	116,281
1860	0.5606	0.5603	+ 3	9	+314	98,596
Total				91,233		2,532,001
Weighted average	0.5292					

---

[1] Figures for column 1, the weighted average and the coefficient of dispersion either given by, or computed according to, W. Lexis, *Abhandlungen zur Theorie der Bevolkerungs-und Moralstatistik*, Jena, 1903, p. 198 *ff.*

given by Lexis. σ_0 was calculated from the values for the coefficient of dispersion Q given by Lexis.

Thus we have the rough coefficient of dispersion

$$Q = 4.3 \text{ (Lexis)},$$

and the corrected coefficient of dispersion

$$Q' = 0.82$$

according to the above method.

$$\sigma_1 = \sqrt{\frac{\Sigma\delta^2}{n-1}} = 10^{-4}\sqrt{\frac{2{,}532{,}001}{19}} = 365.06 \times 10^{-4}$$

$$\sigma_2 = \sqrt{\frac{\Sigma\lambda^2}{n-1}} = 10^{-4}\sqrt{\frac{91{,}333}{19}} = 69.333 \times 10^{-4}$$

According to Lexis' computation $Q = 4.3$,

therefore

$$\sigma_0 = \frac{\sigma_1}{Q} = \frac{365.06}{4.3} \times 10^{-4} = 84.898 \times 10^{-4}$$

$$Q' = \frac{\sigma_2}{\sigma_0} = \frac{69.333}{84.898} = 0.82$$

The probable error according to the Bortkiewicz Formula $E_Q = .6745/\sqrt{2n}$ would be 0.11 so that Q' lies near the lower E_Q limits, $1 \pm .11$, and well within its fourfold reach. As there is no visible reason for subnormal dispersion, we may assume the deviation of Q' from unity is caused by chance and infer that the strong supernormal dispersion of the first way of computing Q has been changed to normal dispersion by eliminating the temporal trend in the change of the sex ratio of deaths of that age and time.

A second similar example was computed from the proportion of recruits of height 166 to 170 cm among all 21-year-old recruits in the Vienna military territory from 1892 to 1912. (See figures in the author's *Grundriss der Statistik* I, 2nd ed., Vienna, 1947, p. 52.) The rough coefficient of dispersion according to Lexis was $Q = 3.251$, the corrected one $Q' = 2.681$, lower than the rough one as expected, but by far not lower so much as in the first example. We conclude that the recruits of the several years were far less homogeneous in their height distribution, so that we failed to obtain normal dispersion as in example 1, after eliminating changes in time caused by the slow but steady increase of height.

TABLE II

First entrance of mentally defective women into institutions in Switzerland. Age distribution[1]

Year	Age Group						Total
	0-19	20-29	30-39	40-49	50-59	60+	
	1	2	3	4	5	6	7

A. Observed figures

1926	19	26	19	12	8	9	93
1927	21	20	17	14	14	7	93
1929	21	29	24	17	9	12	112
1930	28	36	22	14	10	5	115
1931	35	36	35	10	12	7	135
1932	29	54	35	24	9	8	159
1933	30	42	40	16	7	8	143
1935	23	31	40	26	11	2	133
1936	34	51	41	29	14	11	180
1937	36	33	41	18	17	4	149
Total	276	358	314	180	111	73	1312

B. Relative distribution

1926	20.43	27.96	20.43	12.90	8.60	9.68	100.00
1927	22.59	21.51	18.28	15.05	15.05	7.52	100.00
1929	18.75	25.89	21 43	15.18	8.04	10.71	100.00
1930	24.35	31.30	19.13	12.17	8.70	4.35	100.00
1931	25.93	26.67	25.93	7.41	8.88	5.18	100.00
1932	18.24	33.96	22.02	15.09	5.66	5.03	100.00
1933	20.98	29.37	27.97	11.19	4.90	5.59	100.00
1935	17.29	23.31	30.08	19.55	8.27	1.50	100.00
1936	18.89	28.33	22.78	16.11	7.78	6.11	100.00
1937	24.16	22.15	27.52	12.08	11.41	2.68	100.00
Average	21.04	27.29	23.93	13.72	8.46	5.56	100.00

C. Some other values

Q	0.831	1.033	1.094	1.093	1.177	1.428	

---

[1] Figures from *Statistisches Jahrbuch fur die Schweiz* for the respective years.

TABLE II, *Continued*

Year	Arithmetic mean[2]	$\lambda = (A_X - 32.80)$	$\lambda_0 = \frac{\lambda}{Q_A}$	$\lvert\lambda - \lambda_0\rvert$
	1	2	3	4
1926	34.03	+1.23	+1.33	0.90
1927	35.11	+2.31	+0.62	1.69
1929	35.00	+2.20	+0.59	1.61
1930	31.26	−1.54	−0.41	1.13
1931	31.22	−1.58	−0.43	1.15
1932	32.11	−0.69	−0.19	0.50
1933	31.64	−1.16	−0.31	0.85
1935	33.27	+0.47	+0.13	0.37
1936	33.39	+0.59	+0.16	0.43
1937	32.25	−0.55	−0.15	0.40

$$A_X = 32.80; \qquad \sigma_A = \sqrt{\frac{\Sigma\lambda^2}{n-1}} = 1.463;$$

$$\sigma_X = 14.29; \qquad \sigma_{A,0} = \frac{\sigma_X}{\sqrt{N-1}} = 0.394;$$

$$Q_A = \frac{\sigma_A}{\sigma_{A,0}} = \frac{1.463}{0.394} = 3.713$$

---

[2] In computing the arithmetic mean the first and the last groups have been conceived as stretching over 10 years of age.

A third example in Table II concerns a full age distribution, the first entrance of mentally defective women into institutions in Switzerland. The total of 1312 cases is not very great but great enough to allow it to be roughly assumed as a universe from which the samples of the several years are taken.

Here, too, we may apply the above method in the various columns representing the various ages as well as in the column of the arithmetic means. Thus, *e.g.*, the coefficient of dispersion for age 60 is lessened from $Q = 1.428$ to $Q' = 1.145$; the coefficient of dispersion of the arithmetic means $Q_A = 3.713$ to $Q'_A = 1.669$, in the first case using a straight line with the origin in 1932, $y' = 5.6086 - 0.5661x$, in the second case an algebraic curve of the fourth degree $y' = 31.314 + 0.058763x + 0.337451x^2 - 0.015197x^3 - 0.009579x^4$, in graduating the curve of values according to the method of least squares. Using the Bortkiewicz formula for the significance of normality of dispersion $\sigma_Q = 1/\sqrt{2_n}$ in the form of the probable error $E_Q = .6745\ \sigma_Q$, we have the limits $1 \pm .15$, so that the value obtained for Q' for age 60 hints of normality in a much more definite way than in its original form, whereas Q'_A, though considerably diminished, remains supernormal, outside the limits of $4E_Q$.

From those two results we would be inclined to infer that the age distribution observed has less stability as a whole than in its details. I think that inference would be erroneous, as the seeming stability in the age groups is a mere formal consequence of the smallness of the samples, in which case, as we know, the "combinatoric" component (Lexis) of the coefficient of dispersion outweighs the "physical" one, thus giving a misleading picture of normal dispersion.

That assumption seems to be justified by the run of the coefficients of dispersion in all age groups. Here we have the surprising picture of a rising series of coefficients of dispersion, with only one exception, so regular that we may rightly assume those differences to be real. In the youngest group the dispersion is subnormal, in the oldest groups, supernormal; in the middle groups there is uncertainty because of the spurious normality mentioned. Supernormality should not be surprising in view of the fluctuations observed in the arithmetic means; subnormality, as there is no reason to assume negative correlations of the units, could only be explained by regulations influencing the frequency of entrance into lunatic asylums in the youngest group, either in the nature of the illness, or in the policy of admittance in the Swiss lunatic asylums, or in something else which the medical experts may find out. Only one question may be answered by the statistician, whether or not formal causes may have been at work. Such a cause may be the shifts which this and all other age groups are bound to have undergone by the shifts in the whole age distribution indicated in the oscillations of the arithmetic means. That question may be answered in the following way:

It is necessary to shift back the cases causing changes in the age groups as a consequence of that general shift, to an order that would be expected if only random oscillations had influenced that distribution. We may assume that shift parallel throughout the groups, which assumption, though a rough one, is suggested by its simplicity as well as by the difficulty of a clear separation of the significant and the random oscillations in those details.

The way to execute that idea is this: First we compute the coefficient of dispersion of the arithmetic means by the Lexis method, which has already been done above; then we correct the given deviations of the several arithmetic means from that of the sum by dividing them by that coefficient of dispersion Q_A by which division the deviations are transformed to values λ_0 corresponding to $Q_A = 1$. According to those changes in the deviations we establish the corresponding shifts in the arithmetic means and in the whole age distribution, the latter in the way that each group cedes the number of persons to the neighboring group corresponding to the stretch of shift in an equal distribution. The results of that computation are given in Table III.

The coefficients of dispersion Q'' have not considerably changed in the middle columns; the coefficient of the highest age group has diminished towards unity in a way similar to the result of the method of corrected deviations. The coefficient of dispersion in the lowest group, however, has proved to be still subnormal after that adaptation of the distributions. That leads us to assume that that subnormality really existed.

SUMMARY AND CONCLUSIONS

Applying a modification of Lexis' method in taking the deviations of the observed values not from the average of our observations but from their graduated curve, we have attempted to exclude changes in the basic probability of our universe arising in the course of time. Thus we obtain the dispersion which we may assume in an actual case without taking into account changes in the past and possible similar changes in the future. In this way we have made the given observations conform more nearly to the model of samples taken at random under the exacting conditions of the theory of chance. Where then a dispersion resulted greater or smaller than the normal one, we should assume the working of the remaining causes of disturbance: in the case of supernormal dispersion bad mixture in our "bag" from which the samples were taken, or positive correlation between the units of our sample; in the case of subnormal dispersion the working of limiting regulations or negative correlation between the units. Only where normal dispersion was also found, was the application of the limits of the theory of chance verbally right. In the cases of subnormal dispersion it was a superfluous precaution. In the case of supernormal dispersion it was a hazardous trusting in insufficient limits.

Testing all available data according to the method of excluding temporal changes we could determine for what kinds of data those limits are reliable, and for which we have to mistrust them.

In our three examples we have found in one case, the sex distribution of deaths of a higher age, a change from the alleged supernormal into normal dispersion; in two other cases, the proportions of recruits of age 21, 166 to 170 cm. high, and the average age of mentally defective women entering institutions, the dispersion which remained supernormal. In the two latter cases we believed the supernormal dispersion was caused by *insufficient mixture* in the universe, which seems to be the general fault with actual samples, apart from those changes in time excluded by our method.

It seems already apparent from this short study that the sphere of application of those limits is broader than the strict lines of thought admits, but narrower than the liberal line of thought assumes. Therefore, *using the known limits of significance in all cases indiscriminately,* *we may take them as a symbolic reminder that the given results are not as definite as they present themselves, rather than as a real, scientifically proved fact.*

TABLE III

Adjusted figures of Table II

Year	Age group						Total
	0–19	20–29	30–39	40–49	50–59	60+	
	1	2	3	4	5	6	7
A. Adjusted observed figures							
1926	21	26	18	12	8	8	93
1927	24	20	16	14	13	6	93
1929	26	28	23	15	10	10	112
1930	25	35	24	14	11	6	115
1931	31	36	35	13	12	8	135
1932	28	52	36	25	10	8	159
1933	27	41	41	18	7	9	143
1935	24	31	40	25	11	2	133
1936	36	51	40	29	13	11	180
1937	35	33	40	19	17	5	149
Total	277	353	313	184	112	73	1312
B. Adjusted Relative Distribution							
1926	22.58	27.97	19.35	12.90	8.60	8.60	100.00
1927	25.81	21.51	17.20	15.05	13.98	6.45	100.00
1929	23.21	25.00	20.54	13.39	8.93	8.93	100.00
1930	21.74	30.43	20.87	12.17	9.57	5.22	100.00
1931	22.96	26.66	25.93	9.63	8.89	5.93	100.00
1932	17.61	32.70	22.64	15.73	6.29	5.03	100.00
1933	18.88	28.67	28.67	12.59	4.90	6.29	100.00
1935	18.05	23.31	30.08	18.80	8.26	1.50	100.00
1936	20.00	28.33	22.22	16.12	7.22	6.11	100.00
1937	23.49	22.15	26.84	12.75	11.41	3.36	100.00
Average	21.11	26.91	23.86	14.02	8.54	5.56	100.00
Q''	0.766	0.948	1.148	0.845	1.156	1.099	

Résumé

Par une modification de la méthode de *Lexis*, nous avons tâché d'éliminer les changements dans le temps de la probabilité de base. Cette modification consiste à prendre les déviations des valeurs observées non de la moyenne des observations mais de leur courbe ajustée. Nous

obtenons ainsi la dispersion à un moment donné sans qu'elle soit influencé par des changements dans le passé et d'éventuels changements similaires dans l'avenir. Nous avons ainsi rapproché les observations au modèle d'échantillonnage au hasard sous les conditions imposées par le calcul des probabilités. Là où une dispersion est plus grande ou plus petite que la dispersion normale, nous pouvons l'attribuer à l'influence de causes perturbatrices: dans le cas de dispersion hypernormale, que le contenu de l'urne d'où les échantillons sont tirés n'est pas homogène, ou qu'il y a corrélation positive entre les tirages successifs; dans le cas de dispersion hyponormale, l'action de règles restrictives ou dec corrélstion négative entre les tirages. Seulement là où on obtenait la dispersion normale, l'application des limites de la théorie des probabilités était littéralement exacte. Dans le cas de dispersion hyponormale c'était une précaution superflue. Dans le cas de dispersion hypernormale c'était une confiance hasardeuse dans des limites insuffisantes.

En analysant les données suivant la méthode d'élimination des changements dans le temps, l'auteur pouvait déterminer les cas dans lesquels ces limites sont sûres, et ceux dans lesquels il ne faut pas se fier.

Dans les trois exemples donnés par l'auteur, il a trouvé dans un cas, la distribution par sexe des décès à un âge élevé, le changement d'une dispersion considérée comme hypernormale, dans une dispersion normale; dans deux autres cas, la proportion des recrues de 21 ans mesurant 166 à 170 cm., et l'âge moyen de femmes à déficience mentale admises dans des institutions, une dispersion qui reste hypernormale. Dans les deux derniers cas, nous croyons que la dispersion hypernormale est causée par le fait que la population n'est pas homogène, ce qui semble être le défaut général d'échantillonnages, en dessus des changements dans le temps éliminés par la méthode proposée ici.

Cette courte étude montre déjà que la sphère d'application des limites de signification telles qu'elles sont couramment employées est d'une part plus large que certains auteurs l'admettent mais d'autre part plus étroite que ne l'admettent d'autres auteurs. C'est pourquoi, en utilisant ces limites dans tous les cas sans discrimination, on peut les considérer comme un rappel que les résultats présentent un certain degré d'incertitude, plutôt que comme donnant une mesure exacte de cette incertitude.

Discussion

Mr. Bernstein compared Mr. Winkler's treatment of time series with that of Lexis. Mr. Bernstein indicated that Lexis took the moving average into account and calculated a somewhat different dispersion.

Mr. Rasch pointed out that the Lexis coefficient should be interpreted as having two aims. One of these, he said, is to test the hypothesis that the probability in question is a constant one, in which case no adjustment should be made according to time. The hypothesis is wrong if the Lexis coefficient does not fall within the limits given by the chi-square distribution. When the hypothesis is wrong, Mr. Rasch added, an alternative hypothesis may be formulated in such a way that the Lexis coefficient could measure which way the probabilities are varying.

STATISTICAL METHODOLOGY SECTION

*Wednesday, September 10, at 9:30 a.m.**

Uses of Probability Theory in Economic and Social Sciences.

CHAIRMAN :

Tjalling C. Koopmans

POSSIBILITIES AND LIMITATIONS OF THE APPLICATION OF MATHEMATICS AND, PROBABILITY THEORY IN PARTICULAR, IN THE FIELD OF SOCIAL AND ECONOMIC PHENOMENA

by Maurice Fréchet
Paris Faculty of Sciences (France)

Summary

At the request of the Bureau of the International Statistical Institute the author established an inquiry among scientists of different countries, asking their opinion about the possibilities and limitations of the application of mathematics, and probability theory in particular, in the field of social and economic phenomena.

A report on this inquiry, in French, including the full text in the original language of the replies received from 17 scientists, has been published in the *Revue* of the International Statistical Institute.[1]

In this summary we will only formulate opinions and judgments based on the answers to our inquiry.

Some people think that human sciences deal with too complex phenomena for mathematics to be able to describe them. In fact, the primary

---

* Joint meeting with The Economic Society.

[1] Maurice Fréchet, "Dégager les possibilités et les limites de l'application des sciences mathématiques (et en particulier du calcul des probabilités) à l'étude des phénomènes économiques et sociaux (*Revue* Inst. Int. de Stat., 1946, p. 16-51).

elements of mathematics have always been utilized by economists as soon as economics existed. Furthermore, the higher parts of mathematics which have been employed later by a few economists have exerted an indirect influence on the other ones. For mathematics require more precise definitions, they put more emphasis on functional dependence than on causality, and so on. Thus the question really reduces to: are higher chapters of mathematics directly useful to human sciences?

We have considered three main ways to apply mathematics in human sciences. In the first, schemes of very general scope are built on a few postulates in such a manner that they may be expressed mathematically, and that they reduce the economic problem to a mathematical problem. It is this first method which has evoked the most violent criticism because it is generally impossible to verify numerically either its premises or its conclusions. However, this method gives results which may be verified qualitatively, and it helps understanding the ecomomic connections: Similarly when in theoretical mechanics a proposition is "friction being neglected," then in many practical mechanical problems it cannot give numerically good results but it gives a first preliminary idea of what may happen. It is this method which has been brilliantly utilized by Cournot, Walras, Jevons, Dupuit, Pareto, Marshall, Irving Fisher, and so on. It has brought to the foreground the now classical notions of marginal utility, of elasticity, of economic equilibrium, among many other ones.

The second way to utilize mathematics in social sciences appeared later, because it makes great use of statistical data and these data were first inexistent or unreliable. It does not aim at such general description of the economic or social world as the first method but limits itself to the study of particular economic or social phenomena where numerical verification is possible. Such is, for instance, the question whether economic cycles exist, which are their periods if any, and how can they be explained rationally. Such also is the introduction of index numbers, their use and their theory.

The third way to apply mathematics consists in considering that, when there seem to exist functional relations between some economic or social magnitudes, these relations are really random ones or as is said now they are stochastic relations. Instead of trying to find mathematical exact relations between these magnitudes, the question is to find the probability law which governs the set of these magnitudes. This appears the best way for escaping the difficulty presented by the discrepancy between observed magnitudes and magnitudes conputed on the basis of theory.

Studying the limits to be imposed on the application of mathematics to social sciences, there is a general consensus that even the most recent or difficult chapters of mathematics may be legitimately employed in social sciences provided it be understood that their use would not entitle their economic conclusions to be safer.

A real limit is in the limited power of mathematics themselves. When an economic problem has been reduced to a mathematical one, there is not always a mathematical method of solving this last problem.

Some mathematical economists think that mathematical economics is but one of the ways by which economic problems must be approached. Finally, a very important limitation arises from the fact that as soon as an economic law were proved, people hearing of it and trying to avoid its disadvantageous consequences would make it false.

Discussion

Mr. J. R. N. Stone gave two examples of the use of mathematical methods in economic investigations and of the problems to which they are properly applied. The first example, related to the field of pure economic theory, was the work by von Neumann and Morgenstern on the theory of games. The importance of their work said Mr. Stone, is in their pointing out that where imperfect competition exists one does not have the type of mathematical problem which has normally formed the foundation of pure mathematical economics. There is a need for further development in mathematics to express realistically this situation in the economic world.

Mr. Stone's second example, in the field of econometric applications was concerned with the question of model-building. He pointed out that realistic economic models will contain stochastic elements, which introduce an element of unpredictability into the system. He added that it is impossible to foresee all the disturbances to which an economic system may be submitted; moreover, the structure of the system may change. Prediction consists partly in telling what will happen if the structure is changed in a particular way. As a matter of methodology, Mr. Stone believes, all possible factors should be built into economic models.

Mr. Stone summarized his views on the scope of mathematics in economics, stating that the complexity of the situation would probably require exceedingly complicated methods for more than an approximate analysis. He added, on the other hand, that imperfect analyses (such as regression analysis) can be of value to politicians, civil servants, and business men of the present time.

Mr. Rubin remarked to Mr. Stone that mathematical models cannot be improved unless there is information about the possible disturbances (other than there being stochastic relationships) that can be accommodated in the model.

Mr. Wold commented on applications of statistics in experimental fields, as different from those in non-experimental fields. He noted the common experience that different statistical methods are applied in different situations.

Mr. Wold outlined the broad categories of statistical methods: Regarding the factors under analysis as variables in the mathematical or in the probabilistic sense, these variables are observed in the empirical sciences. In causal analysis, one category is concerned with theory working just in terms of the observed variables; in another category it may be found convenient to introduce also other, unobserved variables. In the first category, two cases may be distinguished, where experiments are possible and where they are not. In the second category, where causal factors are unobserved, experiments are in general impossible.

Where experiment is possible, Mr. Wold concluded, the results are reliable; in other fields, however, the analysis is liable to errors of method—problems become more difficult and the results more vague and unreliable.

Mr. Bernstein remarked on the development of mathematics for application to such phenomena as sunspots, and on the developments of mathematics superseding correlation analysis for the analysis of time-series.

Mr. Bernstein also illustrated the problem of mathematical models with his experience in the analysis of cotton prices.

Mr. Rueff stated, with reference to the comments by Messrs. Wold and Bernstein, that the random character of a variable depends essentially on the technical means at the disposal of the observer (or of the government), and on the skill and character of the observer. Moreover, this rondom character of a variable varies with time. He observed that in this respect the social sciences are only in degree in a different position from the physical sciences.

Mr. Lutfalla noted the limitation which exists on the application of mathematics in other fields of social science than economics. If economic theory may be defined as a logic of judgment of economic values, he asked, does there exist also a logic of other social values?

Mr. Fréchet illustrated Mr. Stone's comments with the following examples of the development of mathematics: (1) In theoretical mechanics, "integral differential equations" have been introduced, which permit the study of irreversible phenomena; (2) Mr. Allais has treated mathematically the Pareto definition of maximum satisfaction, dealing with the comparison of two economic states, where one may or may not be able to determine which is preferable.

Mr. Fréchet added that in the mathematical treatment of random phenomena, a distinction should be made between those resulting from the accumulation of many small causes and those which are the effect of one phenomenon of considerable intensity.

Mr. Nanda and Mr. Shastri also participated in the discussion.

STATISTICAL INFERENCE FROM NON-EXPERIMENTAL OBSERVATION: AN ECONOMIC EXAMPLE*

by Jacob Marschak

Cowles Commission for Research in Economics
University of Chicago (United States)

Suppose we want to predict the extent of shortage (or surplus) which will occur if the government fixes the price of a commodity at a certain level. At this price level, supply of the commodity may be equal to demand, but it may also exceed or fall short of demand. When I say that we want to predict the extent of shortage, or of surplus, of the commodity *i.e.*, the difference between demand and supply, resulting from fixing the price at a given level, I use the term prediction in a sense different from that in which Mr. Stone used the term a few minutes ago. When discussing M. Fréchet's report on the results of the questionnaire on the scope of probability theory in social science, Mr. Stone said that, because of the stochastic character of economics, economic prediction is not possible. That is true under a narrow definition of prediction, a definition which if applied rigorously, would exclude prediction from almost all empirical sciences, inasmuch as almost all empirical laws are, strictly speaking, stochastic relations similar to the economic relations that I shall treat in my example.

Let us define *prediction* as the estimation of the probability distribution of a certain random variable or of the joint probability distribution of a set of random variables (the "predictand" set), given the values of certain other variables (the "predictor" set). In that sense, the shortage which we shall denote by y, can be predicted if we can estimate its conditional distribution, given p^*, the price fixed by the government. The best known, though not the only possible technique of estimating any distribution is the "parametric" estimation: one or more parameters of the distribution are estimated on the basis of statistical data, after having used some previous knowledge independent of those data, to specify in advance other parameters, and the form itself, of the distribution function. Similarly we can speak of the *parametric estimation of a stochastic relation* such as the relation

$$f(y, p^*) = v, \tag{1}$$

where v is a non-observable random "disturbance" with zero-mean.

---

* This article will be included in Cowles Commission Papers, new series, no. 32.

The form of the distribution function of v, as well as the form of the function f relating v to the observable variables y, p^*, have to be specified *a priori* before proceeding to the estimation of the parameters of these functions on the basis of observations. A particular case of (1) is the *prediction equation* of y for a given p^*, such as

$$y = R(p^*) + v, \tag{2}$$

where v is a "prediction error", distributed normally with zero mean and with unknown variance σ_{vv} while $R(\cdot)$, a "regression function," is specified with respect to its form (*e.g.*, linear) but not with respect to all its parameters. It follows that, for every fixed value of p^*, the predictand y will have a normal distribution characterized by the expectation (mean) of y given p^*,

$$\mathcal{E}(y \mid p^*) = R(p^*), \tag{3}$$

and by the variance of y given p^*, *i.e.*, the square of "the standard error of the estimate,"

$$\sigma^2 (y \mid p^*) = \sigma_{vv}. \tag{4}$$

Hence the estimation of the unknown parameters of the regression function $R(\ )$, and of the unknown parameter σ_{vv} of the distribution of v, will suffice to estimate the stochastic "prediction relation" (2), and also to predict the variable y.

I shall now try to show that useful prediction requires different procedures according to whether experiments are or are not possible. Economics will be shown to require techniques not developed in experimental sciences. Specifically: the economist may have to predict the results of introducing price control, not on the basis of experiments with price control but on the basis of observations which were made in the absence of price control.

Denote by x^d_t the demand, and by x^s_t the supply, at time t. Hence, shortage $y_t = x^d_t - x^s_t$. Let p_t be the price which is either the free market price or the government fixed price as the case may be. Suppose that for any t, the behavior of buyers and that of sellers is described, respectively, by

$$x^d_t = u^d_t + L(p_t, z_t; \alpha^d)\text{: demand equation; and} \tag{5}$$

$$x^s_t = u^s_t + L(p_t, z_t; \alpha^s)\text{: supply equation.} \tag{6}$$

Here both demand and supply depend on price and, in addition, on a whole set of "*predetermined*" *variables* such as weather, tax-rates, past crops, etc. Some of these variables may appear in both equations, others in only one of the two equations. We denote the set of all predetermined given variables by a single symbol z_t. Both the demand and the supply functions are linear, as indicated by the symbol $L(\ )$; but they have different and, in general, unknown coefficient-sets, α^d and α^s respectively. For brevity, we shall also use a single symbol $\alpha \equiv (\alpha^d, \alpha^s)$. Finally, the behavior of both sellers and buyers is stochastic: we suppose both to be subject to "random shifts," or disturbances, u_t^d, u_t^s. These are non-observable random variances and covariances forming a matrix

$$\sigma \equiv \begin{Vmatrix} \sigma^{dd} & \sigma^{ds} \\ \sigma^{ds} & \sigma^{ss} \end{Vmatrix}.$$

Here σ^{ds} need not be zero since buyers' and sellers' behavior shifts may be correlated. The average "violence" of these disturbances is measured by σ^{dd}, σ^{ss}, respectively.

The parameters α of the demand and supply functions are properties of behavior or technology; so are the distribution parameters σ. All these parameters are necessary to describe the "structure" of the market. They are, however, not sufficient. The stochastic system (5), (6) consists of only two equations in the three variables (x_t^d, x_t^s, p_t) that are not determined by outside forces. Therefore, the values of these variables are not yet determined. The system becomes complete if supplemented by a third equation. Under price-control, this equation is

$$p_t = p_t^*, \tag{7'}$$

where p_t^* is some constant chosen by the government for the time t. In a free market we may have, instead,

$$x_t^s = x_t^d \tag{7''}$$

if we assume that sellers adjust production instantaneously to wipe out the shortage or surplus, shown by decreasing or increasing stocks on hand—the simple static theory of economic textbooks which at best is a first approximation. (A somewhat more realistic, dynamic theory might introduce a positive lag λ, as in $x_{t+\lambda}^s = x^d$; or in $p_{t+\lambda} - p_t = k(x_t^d - x_t^s)$, $k > 0$.)

We can combine equations (7') and (7'') into a single one,

$$(x_t^d - x_t^s)\delta + (p_t - p_t^*)\epsilon = 0, \tag{7}$$

$$\text{where } (\delta, \epsilon) \begin{cases} = (1, 0) \text{ in free market;} \\ = (0, 1) \text{ under price control,} \end{cases}$$

The "structure" of the market is then described by the numerical values of the parameters $(\alpha, \sigma, \delta, \varepsilon)$. We see that structure depends on men's behavior, technology, legal rules.

Now let me proceed to our problem of predicting the extent of shortage, $y_t = x_t^d - x_t^s$, as the result of fixing the price at $p_t = p_t^*$. Usually, observations will have been collected under the condition that $(\delta, \varepsilon) = (1,0)$. But prediction is to be made for the situation $(\delta, \varepsilon) = (0, 1)$. There is a change of structure—a *known* change to be sure—between the time of observed and the time of predicted events.

This is characteristic of the non-experimental case. I shall call it Case B. When experiments are possible I shall speak of Case A. In our economic problem, an imaginable experiment would consist in deliberately fixing the price at various values and observing the results. I shall briefly study this imaginable Case A before proceeding to the actual, non-experimental Case B.

Let us, then, assume that we can fix prices, from experiment to experiment, at varying values $p_1^*, p_2^*, \ldots p_T^*$. That is $(\delta, \varepsilon) = (0,1)$ for $t = 1, \ldots, T$. Our object is to predict the shortage, say, $y_{T+1} = x_{T+1}^d - x_{T+1}^s$, when, at time $T+1$, the price is fixed at p_{T+1}^*. This prediction is obtained by at least two, logically distinct procedures. Both occur in experimental sciences. I shall call the two procedures, the *specific experiment* and the *general experiment*, and denote them, respectively, by A1 and A2. Examples of specific experiment are the testing of model bridges, or the testing of airplanes in aerodynamic tunnels. Here the experiment consists in reproducing the specific structural change whose effect one wants to predict. The original structure is not the subject of investigation. General experiments, on the other hand, are designed, not to predict the results of a single specific structural change, but to make possible the prediction of the results of any possible structural change. Experiments made in physical laboratories to establish fundamental laws of physics or chemistry are of this kind. The time required to boil an egg can be estimated by boiling eggs; but, in principle, it can also be estimated by studying first the general properties of protein molecules, in which case not only the answer of the egg-boiling question, but of many other questions as well, will be prepared.

In our economic case, a *specific experiment*, (A1) would run as follows. The experiment—conducted, say, in a small community—replaces the random variables p_t in (5), (6) by constants p_t^*, for t running from 1 through $T+1$. Therefore, if we define

(8) $$v_t = u_t^d - u_t^s \text{ and}$$

(9) $$\beta = \alpha^d - \alpha^s, \text{ we have}$$

(10) $$y_t = v_t + L(p_t^*, z_t; \beta), \qquad t = 1, \ldots, T+1,$$

Since v_t is, by (8), normally distributed with zero-mean, equation (10) is a prediction equation, exactly like (2). As in (3), (4), the conditional mean and conditional variance of the predictand y_t, are functions of given values of the predictor set (p_t^*, z_t):

$$(11) \qquad \mu_{y_t} = \mathcal{E}(y_t \mid p_t^*, z_t) = L(p_t^*, z_t; \beta)$$

$$(12) \qquad \sigma_{yy} = \mathcal{E}\left\{ (y_t - \mu_{y_t})^2 \,\middle|\, p_t^*, z_t \right\} = \sigma_{vv} = \sigma^{dd} + \sigma^{ss} - 2\sigma^{ds}, \text{ by } (8).$$

We see that the distribution parameters μ_{y_t}, σ_{yy} of the predictand are obtained *without estimating the structural parameters* $\bar{\alpha}$, σ. Instead we obtain unbiased and efficient estimates of β and σ_{vv} directly from the observations $(y_t, p_t^*, z_t; t = 1, \ldots, T)$ by minimizing $\sum_1^T v_t^2$ with respect to β and σ_{vv}: (the method of least squares). We can thus estimate, for given p_{T+1}^* and z_{T+1}, the conditional mean of y_{T+1} by (11); and the conditional variance $\sigma_{yy} = \sigma_{vv}$. That is, we can predict y_{T+1}, without estimating the structural parameters α, σ.

We can also use our economic example to illustrate the experimental procedure A2, the use of *general experiment*. Structural parameters are estimated; from these estimates, the effect of any given structural change is predicted as the need arises. The experiment itself will again consist (in our example) in replacing the random variables p_t produced in the free market, by fixed constants p_t^* $(t = 1, \ldots, T)$. But this time we measure not the shortages y_t directly, but, separately, the demand x_t^d and the supply x_t^s. The relation between the single random variable x^d (or x^s) and the predetermined variables p^*, z is again like that in (2). We can therefore estimate the structural parameters α^d, α^s by two least square regressions of, respectively, x^d and x^s on the predetermined variables p^*, z; and we estimate the structural parameters σ^{dd}, σ^{ds} from the regression residuals. To predict the shortage y_{T+1} for given p_{T+1}^*, z_{T+1}, we obtain the parameters β and σ_{vv}, and hence the parameters μy_{T+1}, σ_{yy}, by (9), (11), (12): that is, we are able to predict the effect of the transformation of the original free market structure $(\alpha, \sigma, 1, 0,)$ into the controlled market structure $(\alpha, \sigma, 0, 1)$. But in addition, we can now also predict the result of any other given change undergone by the original free market structure: for example to predict what happens if a technological change raises some of the coefficients α^s of the supply equation by a given percentage; or if an expected change in tastes or in population raises in a given proportion some of the demand coefficients α^d, or, possibly the violence of random demand shifts σ^{dd}, etc.

Consider now the Case B: the non-experimental situation, usual in economics. In this case, no observations are available except those made in the free market, with prices p_t determined by supply and demand, and therefore depending on their random shifts u_t^d, u_t^s. Thus for $t = 1$, ..., T, the structure is $(\alpha, \sigma, 1, 0)$. But for $t = T+1$, the structure is $(\alpha, \sigma, 0, 1)$. Equation (10) is valid for $t = T+1$, but not for earlier time points. Estimates of β, σ_{vv} necessary, as before, to predict y_{T+1} from (10) cannot be obtained from past observations on (y_t, p_t^*, z_t) since, for $t = 1, ..., T$, we have $y_t = 0$, and p_t^* is not defined. During this past period, price p_t, was not fixed. Instead, it was a random variable, having a joint distribution with the variable x_t (by which we shall now denote both demand and supply since both were equal); this joint distribution, conditional upon the values of the predetermined variables is, under the conditions of the problem, again normal. Denote by γ the parameters of this joint distribution of the observables x_t, p_t (that is, the set γ consists of the conditional means and variances and covariance of x_t and p_t.

Since p_t, x_t depend on u_t^d, u_t^s, the parameters γ are uniquely determined by the structural parameters σ, α. If the converse were always true, the structural parameters α, σ could always be determined from the knowledge of γ; while the parameters γ of the joint distribution of p_t, x_t can, of course, always be estimated from observations on these variables. Whether, in a particular case, the parameter sets α, σ (or some elements of them, *e.g.*, the coefficients α^d of the demand function) are in fact uniquely determined by γ, depends on the assumed economic theory, or "model". For example, if in the model (5), (6), the set of predetermined variables z is represented by a single variable (say, income tax-rate z') in the demand equation, and by another single variable (say, rainfall z'') in the supply equation, then it can be shown that only one structure can be compatible with a given distribution of x, p. In such a case we say that the model makes the whole structure identifiable. In some other case the model might make only a part of the structure—such as the demand coefficients α^d—identifiable; or no single parameters might be identifiable. This does not depend on the number of observations in hand but entirely on the model which the economist has assumed on the basis of some previous knowledge. The conditions of identifiability have been given by Koopmans and Rubin for an important class of cases. The concept itself was defined by Koopmans and has been further studied by L. Hurwicz (all of the Cowles Commission). In a less rigorous way economists have been aware of the problem for a long time but more especially since Haavelmo pointed it out in his book *Probability Approach to Econometrics*.

Suppose then that the chosen model is such as to make all the parameters of our structure identifiable. Then α and σ can be derived from γ; but γ can always be estimated from the observations on p^*, x, being the set of parameters of their distribution, hence α and σ can be estimated. Therefore, β and σ_{vv} can be estimated, because of (9), (12); and the shortage y can be predicted, since its conditional mean and its variance are determined by (11), (12).

Knowledge of the structure can help not only in predicting the result of the structural change just considered, *viz.*, the introduction of price control; it can be also used to predict the result of any other structural change, provided the change can be spelled out as a well-defined transformation of the original structure (α, σ, 1, 0) which prevailed when the observations were made. Thus, in our non-experimental Case B as well as in the case of general experiments A2, prediction is preceded by structural estimation. Then the knowledge of structure makes it possible to predict the effect, not only of one given structural change (as is the case with specific experiments, A1), but of any well-defined structural change. The estimation of structure, the establishing of a theory, whether with or without the help of experiments is therefore not only a matter of scientific curiosity, it has also the practical value of providing for a number of future contingencies.

The case of general experiments (A2) and the case of non-experimental observations (B) have in common the same aim of estimating the original structure, rather than producing the changed structure by specific experiments (A1). But this aim is reached by different methods. Experiments are common to A1 and A2. In the present example, they call for the application of the method of least squares, which arises as a special case of maximum likelihood estimation. In Case A1 the probability density of y given p^*, z, taken at the observed point $(y_1, \ldots, y_T)$, was maximized with respect to the estimated parameters, and this is equivalent to maximizing $-\sum_1^T v_t^2$ (the exponent of the normal probability density function considered), *i.e.*, to minimizing the sum of squares of the residuals v_t. In Case A2 the probability densities of the independent variables x^d, x^s were similarly maximized, each one separately, In the Case B, maximum likelihood estimates are obtained by maximizing the probability density not of a single variable but of two jointly dependent ones x, p: this is a generalization of the least squares principle. In *Econometrica* 1943, Haavelmo pointed out the joint dependency of variables in economic models, and Mann and Wald studied the properties of maximum likelihood estimates of structural parameters. At the Cowles Commission the method has been developed by Koopmans and Rubin for the estimation of all structural parameters simultaneously. When one or more structural equations are to be estimated separately, more flexible though less efficient procedures are applied, developed at the Cowles Commission by T. W. Anderson, Girshick and Rubin for the case of a single equation, and by Rubin for the case of several equations. All these methods include cases where the jointly dependent variables occur with time lags.

The economist faces the combined disadvantages of the meteorologist and the engineer. At least until artificial rain was produced a few weeks ago, meteorologists were not able to experiment with weather; nor did they need to, as they did not consider it their task to change weather. Engineers, on the other hand, are required to predict the results of their interventions with nature; but in this task they are helped by experiments. The economist is deprived of experiments that change the structure, yet has to predict results of policies consisting in structural change.

This picture is incomplete. Some economic policies consist, not in changing the structure but merely in fixing the values of exogenous variables. Also, structural estimation may call for different methods in economics, where observations cannot be duplicated; and, say, in astronomy where they can. The principles as well as the techniques of structural estimation require much further study.

Résumé

Les objectifs de l'économiste sont analogues à ceux de l'ingénieur, mais les données dont il dispose sont pareilles à celles du météorologue. L'économiste doit souvent estimer les effets d'un changement futur (voulu ou attendu) du mécanisme même dont il tire ses données, sans pouvoir produire aucun de ces changements.

Soit x_t^d = la demande, x_t^s = l'offre, p_t = le prix, à l'instant t. Il s'agit de prédire l'excédent de demande ou d'offre (pénurie ou accumulation des stocks) $y_{T+1} = x_{T+1}^d - x_{T+1}^s$ qui résulterait de la fixation du prix P_{T+1}^* par le gouvernement; la prévision est fondée sur des observations successives aux temps $t = 1, \ldots, T$.

Supposons que pour un instant quelconque t,

$$x_t^d = u_t^d + L(p_t, \bar{z}_t\,;\, \bar{\alpha}^d);\ x_t^s = u_t^s + L(p_t, \bar{z}_t\,;\, \bar{\alpha}^s); \tag{1}$$

$$(x_t^d - x_t^s)\delta + (p_t - p_t{}^*)\varepsilon = 0 \tag{2}$$

(les symboles surlignés représentent des vecteurs ou matrices et L représente une fonction linéaire). Ici, la distribution jointe des "chocs aléatoires" u^d, u^s est normale, indépendante du temps; leurs moyennes sont zéro; leurs écarts quadratiques moyens et la covariation forment la matrice

$$\begin{bmatrix} \sigma_{dd} & \sigma_{ds} \\ \sigma_{ds} & \sigma_{ss} \end{bmatrix} \equiv \bar{\sigma} ;$$

les $\bar{z}$ sont des variables exogènes telles que les conditions atmospheriques ou les taux d'impôts; $(\bar{\alpha}^d, \bar{\alpha}^s) \equiv \bar{\alpha}$ sont les coefficients des fonctions de demande et d'offre; $(\delta, \varepsilon) = (1, 0)$ dans un marché libre, et $= (0, 1)$ quand le prix est fixé. Les valeurs numériques de $(\bar{\alpha}, \bar{\sigma}, \delta, \varepsilon)$ définissent la "*structure*" du marché. Instituer le contrôle des prix veut dire changer la structure de $(\bar{\alpha}, \bar{\sigma}, 1, 0)$ à $(\bar{\alpha}, \bar{\sigma}, 0, 1)$.

Cas A: *les prix peuvent être fixés d'une expérience à l'autre*

à $p_1^*, \ldots, p_T^*$. Dès lors, $(\delta, \varepsilon) = (0, 1)$ pour $t = 1, \ldots, T+1$. Deux façons de procéder sont alors possibles:

Procédé A_1: Les effets d'un changement structurel sont prédits directement sur la base de l'expérience, sans que les paramètres structurels soient estimés. (Cas analogue: avions soumis à l'épreuve dans un tunnel aérodynamique). L'expérience remplace les variables aléatoires p_t dans (1) par les constantes p_t^*. Nous avons

(3) $y = v_t + L(p_T^*, \bar{z}_t; \bar{\beta}),\ t = 1, \ldots, T+1$; où $v_t = u_t^d - u_t^s$ et

(4) $$\bar{\beta} = \bar{\alpha}^d - \bar{\alpha}^s$$

Dès lors, y a une distribution normale (conditionnelle) pour $\bar{z}$ et p^* donnés, ayant comme moyenne

$$\mu_y = E\{y \mid p^*, \bar{z}\},\text{ et écart quadratique moyen}$$

(5) $$\sigma_{yy} = E\{(y-\mu_y)^2 \mid p^*, \bar{z}\} \equiv \sigma_{vv} = \sigma_{dd} + \sigma_{ss} - 2\sigma_{ds}$$

L'estimation de $\bar{\alpha}$ et $\bar{\sigma}$ n'est pas requise! Sur la base des observations $(y_t, p_t^*, z_t;\ t = 1, \ldots, T)$ l'on obtient des estimations "efficaces" des paramètres $\bar{\beta}$, σ_{vv} en minimisant $\sum_1^T v_t^2$ par rapport à ceux-ci (méthode de régression). Sur la base de ces estimations, et pour des p^*_{T+1}, $\bar{z}_{T+1}$ donnés, nous estimons μy_{T+1} et σ_{vv} et pouvons ainsi prédire y_{T+1}.

Procédé A_2: L'expérimentation nous aide à estimer les paramètres structurels $(\bar{\alpha}, \bar{\sigma})$. Sur la base de ceux-ci l'effet d'un changement structurel quelconque peut être prédit. (Analogue: détermination et application subséquente des lois physiques). Puisque l'expérimentation permet de substituer des constantes fixes p_t^* aux variables aléatoires p_t

dans (1), on peut estimer les paramètres $\bar{\alpha}^d$, $\bar{\alpha}^s$ par la régression de x^d et x^s sur p^*, $\bar{z}$; et les paramètres $\bar{\sigma}$, en se basant sur les résidus de régression (c'est-à-dire les écarts entre x^d ou x^s et leurs valeurs estimées). Pour estimer les effets d'un changement structurel dû à l'instauration du con trôle des prix, l'on dériverait les estimations de $\bar{\beta}$ et σ_{vv} de celles de $\bar{\alpha}$, $\bar{\sigma}$ d'après (4) et (5), ce qui permettrait la prévision de y_{T+1}, comme en A_1 ci-dessus. En outre, l'effet d'autres changements structurels, c'est-à-dire d'autres transformations de la structure originelle ($\bar{\alpha}$, $\bar{\sigma}_1$, 0), peut être estimé une fois que la structure originelle elle-même a été estimée. *Cas B*: *L'on ne dispose d'autres observations que celles relatives au marché libre* où les prix $p_1, \ldots, p_T$ sont déterminés par l'offre et la demande. A l'instant $t = T+1$, (3) est encore vrai. Mais les estimations de $\bar{\beta}$, σ_{vv}, nécessaires pour prédire y_{T+1} d'après (3) ne peuvent être obtenues des observations passées de (y_t, p_t^*, $\bar{z}_t$) puisque, quand $t = 1, \ldots, T$, p_t^* n'est pas défini et $y_t = 0$. Au cours de cette période, p_t et x_t^d ($= x_t^s = x$) sont conjointement distribués normalement pour z_t donné. Les paramètres de cette distribution conditionnelle conjointe—appelons les γ—sont des fonctions de $\bar{\alpha}$, $\bar{\sigma}$ seulement, puisque p_t, x_t sont des fonctions de u_t^d, u_t^s, $\bar{\alpha}^d$, $\bar{\alpha}^s$. Inversément, $\bar{\alpha}$, $\bar{\sigma}$ peuvent être ou ne pas être déterminés uniquement par γ. S'ils sont ainsi déterminés, la structure est dite identifiable. $\bar{\alpha}$, $\bar{\sigma}$ peuvent alors être estimés d'après les observations de p_t, x_t ($t = 1, \ldots, T$), puisque γ peut toujours être estimé d'après ces observations. Sur la base des estimations de $\bar{\alpha}$ ($\equiv \bar{\alpha}^d, \bar{\alpha}^s$) et $\bar{\sigma}$ ($\equiv \sigma_{dd}, \sigma_{ss}, \sigma_{ds}$) des estimations de $\bar{\alpha}$ et σ_{vv} sont obtenues par (4) and (5). En outre (comme dans A_2) une fois $\bar{\alpha}$, $\bar{\sigma}$ connus, les effets de changements structurels autres que l'instauration d'un contrôle des prix peuvent être estimés.

Dans le cas B nous devions prédire l'effet d'un changement structurel sans que ce changement puisse faire l'objet d'une expérience. Ce cas écarte le procédé A_2, et requiert l'estimation de la structure originelle. Mais cette estimation ne peut pas non plus être étayée par l'expérimentation comme dans le cas A_2. En appliquant la méthode de régression, en A_1 et A_2, la densité de probabilité d'une variable aléatoire unique telle que y, x^d ou x^s est undue maximum pour des valeurs fixes données de p et $\bar{z}$. Dans B, une généralisation du même principe exige que soit maximum de la densité simultanie des observations de deux variables aleatoires dépendantes x, p, lorsque $\bar{z}$ est fixé. Dans *Econometrica* 1943, ceci fût mis en évidence par Haàvelmo, et Man et Wald y démontrèrent les propriétés des estimations structurelles ainsi obtenues. Cette méthode a été développée à la Commission Cowles, où elle a été appliquée à des systèmes dynamiques comprenant jusqu'à 11 équations.

Discussion

Mr. J. R. N. Stone commented that Mr. Marschak's paper shows forcibly the importance of correct theory as a prerequisite for analytical work with observations. The observations do not determine which of a set of simple transformations of the structure is in fact the fundamental structure of the economy. Only with observations covering several different structural situations can the basic constants be determined out of which structural parameters appropriate to any particular period can be built.

Mr. Stone stated that he understood economic models to consist of equations expressing structural and not necessarily causal relations. Causal relationships may be found by transformation of the system expressing each variable in terms of all the pre-determined variables of the system. Mr. Stone then referred to Mr. Wold's discussion of recursive systems of equations,[1] and indicated that if economic theories comprise such systems the problem of transformation would be less difficult.

Mr. Tinbergen asked Mr. Marschak to give examples of what he meant by identification, showing when it is and when it is not possible.

Mr. Tinbergen also asked whether the change in tax rates, which Mr. Marschak suggested as an example of the value of one of the two to be introduced, might not be considered as a change in price levels.

Mr. Tinbergen's third question referred to Mr. Stone's comments on causal relations, stating that he believed an equation of the reduced form could be called only an *indirect* rather than a *direct* causal relationship. Mr. Tinbergen asked whether significance should be attached to distinction. He illustrated his point by noting the demand equation whose reduced form will usually be given in the form of a sales equation where quantities sold are expressed as a function of demand and supply factors without the introduction of price. Mr. Tinbergen suggested that such an equation did not represent a direct causal relationship as it acts in the minds of people demanding or supplying the commodity.

Mr. Marschak pointed out to Mr. Stone that, regarding the situation where these are observations under two different structures, an example of an experiment may be found in the use of data for prewar and war years; certain parameters have been constant while others have diverged. The problem is then to predict the result of a change in structure resulting from a proposed economic policy change. Mr. Marschak indicated how such a problem might be solved.

---

[1] See these Proceedings, Volume V (The Econometric Society) pp. 1–22.

Mr. Marschak next took up the question of the possibility of a nonformal definition of structural equations, raised by Messrs. Stone and Tinbergen. Fundamental laws, in contrast to formally-defined structural equations (which are always relative to a given policy), are understood by Mr. Marschak to express the behavior of specified social groups, and also certain external circumstances (*e.g.*, legal rules). Mr. Marschak pointed out that there may be more than one system of equations for each social group.

Mr. Marschak observed, in response to the questions on causality, that in the case of causal relations he assumed that the effect usually followed the cause, with some temporal lag, however short. This lag may be ignored if very small, and there is then no way of distinguishing cause from effect. Mr. Marschak thought it not of great importance to determine causal relations, but agreed with Mr. Stone's consideration of reduced forms as causal relationships, especially when lagged variables are included. Mr. Marschak indicated to Mr. Tinbergen that a causal relation need not express a fundamental law.

Mr. Craig stated that he liked to think of economic coordinates as constituting a grid, and discussed the application of an instrument from physics to such a grid: (given a number of conductors in a field, the situation of the structure is given by the capacity and coordinates of each of these conductors. Bringing a unit of electricity on one of the conductors, the distribution of quantity of electricity and of potential is altered in accordance with a set of coefficients. In a grid of prices, an increase in the steel price will have primary effects and also secondary effects through the effect of a change in steel price on the steel price.)?

Mr. Wold added to the discussion of recursive systems, pointing out that they may be used for prediction (1) if future disturbances may be assumed to the zero, and (2) if a future development is given. In the second case, there must be the assumption that the change does not alter the pattern of other parts of the system.

Mr. Hurwicz gave an example showing how identification is achieved. He added that to assure the presence of identification, assumptions are required that may be made on the basis of theory. If the theory is incorrect, results may be vitiated all the way through. Referring back to the topic of Mr. Fréchet's paper Mr. Wold noted that this indicated how lack of an adequate mathematical technique may imply incorrectness of statistical results and forecasts.

Mr. Fréchet observed that Mr. Marschak has distinguished two forms of random events: small ones which are not mentioned in the theory and a single big random event which forces a complete change in the system of equations.

Mr. Fréchet also protested the use of the least squares method.

Mr. Marschak replied to Mr. Fréchet that there must be a further distinction among the small causes between random variables and non-random variables, and suggested to Mr. Tinbergen that it would be difficult to assign a probability law to tax-rates so that they might be a considered random. Mr. Marschak suggested a classification of exogenous variables between the z—observable, and the u—random but non-observable. He noted that there are also non-measurable, non-random variables.

Mr. Marschak suggested that a difficulty would arise in the application of the recursive method, if the time periods become small. In that case, it cannot be assumed that disturbances in two consecutive points of time are non-correlated.

A SHORT-CUT METHOD FOR DISTINGUISHING BETWEEN RIGID AND DISTURBED PERIODICITY

by Herman O. Wold
University of Upsala (Sweden)

1. INTRODUCTORY

When we are dealing with time series which present wave-like oscillations, it is of fundamental importance to find out whether the waves are precisely regular in the sense of an exact, "rigid" periodicity, or whether the periodicity is disturbed by random shifts in the period length. Considering the two cases from the viewpoint of prediction, it is clear that if the waves are of constant length, efficient forecasts of the series may be obtained for any future time point, however remote, while random shifts in the period length will make the series efficiently predictable over short ranges only. In the analysis of rigid periodicity, the main tool is the Schuster *periodogram*. Disturbed periodicity, first studied systematically by G. U. Yule,[1] has in recent years attracted more and more interest in statistical research. A chief tool for the analysis of disturbed periodicity is the *correlogram*, to use the term which I have proposed for the graph of autocorrelation (serial) coefficients.[2]

A new approach to the problem in question will be briefly outlined in the present note. The approach should be regarded as a short-cut method, devised to give preliminary results which might serve to guide the subsequent application of the more toilsome methods of periodogram and correlogram analysis.

As set forth in some detail in § 2, the new device is based on the correlation coefficient r as defined by (1). Under certain simplifying assumptions it is found that the expected value of r_k is $-\frac{1}{2}$ if the given series is strictly periodic, while for sufficiently large k the expected value of r_k is 0 if the periodicity is disturbed. By our device the problem of distinguishing between rigid and disturbed periodicity is thus transformed into the problem of testing two alternative hypotheses about a correlation coefficient. Our test problem is essentially non-parametric; in fact, empirical time series present such a host of widely different patterns that the hypotheses about their structure cannot adequately be brought together into a single parameter system.

---

[1] G. U. Yule, "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers," *Trans. Roy. Soc.*, (A), 226.

[2] H. Wold, *A Study in the Analysis of Stationary Time Series*, Upsala, 1938.

The testing of correlation hypotheses is known to require a fairly large number of observations. Accordingly, we cannot expect the new criticism to be decisive unless the time series dealt with covers a fairly large number of waves. Now this remark is quite in line with earlier experience as to the difficulties of trying to find out whether periodicity is rigid or disturbed.[3]

The material of the present note forms part of a larger project, which includes, *inter alia*, an investigation into the significance levels of the correlation criticism suggested. The sampling analysis is planned to make extensive use of artificially constructed time series (*cf.* illustration C in § 3). The project further includes a study of suitable methods for locating the genuine maxima and minima of a given time series, *i.e.*, genuine as distinct from small, random-like fluctuations.

I take this opportunity to express my gratitude to the Rockefeller Foundation for sponsoring the project mentioned by a generous grant. Dr. T. W. Anderson and Dr. S. Malmquist have read the present note in manuscript and given valuable comments. To both of them I extend my sincere thanks.

2. DESCRIPTION OF THE METHOD

Let $x(t)$ denote the time series under analysis. Recording the time points for which the successive waves in $x(t)$ present maxima, we obtain a sequence which we denote

$$t_0, t_1, \ldots, t_{\nu-1}, t_\nu, \ldots, t_n \qquad (t_{\nu-1} < t_\nu).$$

Now let $d_{\nu k}$ be the distance between one maximum and the kth preceding maximum, in symbols

$$d_{\nu k} = t_\nu - t_{\nu-k} \qquad (\nu = k, k+1, \ldots, n).$$

Thus prepared, let k be fixed, and form the serial correlation coefficient between $d_{\nu k}$ and $d_{\nu+k, k}$ according to the formula

$$r_k = \frac{\sum_{\nu=k}^{n-k} (d_{\nu k} - m_{1k})(d_{\nu+k,k} - m_{2k})}{\sqrt{\sum_{\nu=k}^{n-k} (d_{\nu k} - m_{1k})^2 \sum_{\nu=k}^{n-k} (d_{\nu+k,k} - m_{2k})^2}} \tag{1}$$

where

$$m_{1k} = \frac{1}{n - 2k + 1} \sum_{\nu=k}^{n-k} d_{\nu k}, \qquad m_{2k} = \frac{1}{n - 2k + 1} \sum_{\nu=2k}^{n} d_{\nu k}.$$

We shall now examine r_k for different types of time series $x(t)$.

---

[3] See, *e.g.*, H. Wold, *loc. cit.*, p. 119. *Cf.* also M. G. Kendall, *The Advanced Theory of Statistics, II*, London, 1946, p. 412.

On the one hand, suppose the periodicity of $x(t)$ is disturbed in the sense that $x(t)$ and $x(u)$ are independent in the limit as $t - u \rightarrow \infty$. Fixing k sufficiently large, and writing

$$d_{\nu+k,k} = t_{\nu+k} - t_{\nu} = m_k + \xi_{\nu},$$

it is seen that ξ_ν will essentially be determined by values $x(t)$ over a t-interval which is far away from t_ν. Hence ξ_ν will be approximately independent of

$$d_{\nu k} = t_{\nu} - t_{\nu-k} = m_k - \xi_{\nu-k} \, .$$

It follows that the correlation coefficient r_k as defined by (1) will have an expected value near to zero, more precisely

$$E(r_k) \rightarrow 0 \qquad (k \rightarrow \infty). \tag{2}$$

On the other hand, suppose the waves in $x(t)$ are rigid in the sense that $x(t)$ is additively composed of two components, say

$$x(t) = y(t) + z(t),$$

where $y(t)$ is strictly periodic with period p,

$$y(t + p) = y(p),$$

while $z(t)$ is irregular in the sense that $z(t)$ and $z(u)$ are independent in the limit as $t - u \rightarrow \infty$. On these assumptions, we may write

$$t_\nu = \tau_\nu + \eta_\nu,$$

where

$$\tau_\nu = \tau_0 + \nu.p \qquad (\nu = 0, 1, \ldots, n)$$

are the time points for which $y(t)$ has its maxima, while η_ν is a random variable which represents the deviation in time between an observed maximum and the corresponding maximum of the underlying periodic component. By hypothesis, η_ν is asymptotically independent of $\eta_{\nu+k}$ for large k, and $E(\eta_\nu) = 0$. Now to evaluate $E(r_k)$, we observe that the main term of the numerator in (1) will be

$$E(\eta_\nu - \eta_{\nu-k})(\eta_{\nu+k} - \eta_\nu) = -E(\eta_\nu{}^2),$$

while the denominator gives

$$E(\eta_\nu - \eta_{\nu-k})^2 = E(\eta_{\nu+k} - \eta_\nu)^2 = 2\, E(\eta_\nu{}^2).$$

For the mathematical expectation of r_k we accordingly obtain, as a first approximation,

$$E(r_k) \cong -\tfrac{1}{2} \qquad (k > a/p). \tag{3}$$

In view of (2) and (3), the coefficients r_k may serve as a rough criterion for deciding whether the periodicity of $x(t)$ is rigid or disturbed.

3. ILLUSTRATIONS

A. *Air temperature, Upsala, 1865-1945.* The air temperature at Upsala has a yearly maximum, usually in July. Detailed records of the maximal temperature are available since 1865. Taking from these records the time points t_ν of yearly maxima, and applying formula (1), we obtain

$$r_1 = -0.427,\ r_2 = -0.748,\ r_3 = -0.516.$$

The values approximate —1/2, thus pointing towards the hypothesis of rigid periodicity, as expected. The interpretation is that if the maximal temperature comes early (late) one year, this has no marked effect on the timing of the maximum next year.

B. *Wolfer's sunspot numbers, 1749-1941.* The sunspot intensity presents clear cycles, where the length of the waves varies between 7 and 17 years. In the study already referred to, G. U. Yule chose this phenomenon for illustrating his new statistical methods of investigating disturbed periodicity. Now Wolfer's sunspot numbers present 17 maxima in the period 1749-1945. Applying formula (1) to this material, we obtain

$$r = -0.052.$$

This would be in agreement with the hypothesis of disturbed periodicity. The situation is perhaps not quite clear, however, for if we proceed to the distances between non-adjacent minima, we obtain

$$r_2 = -0.452,\ r_3 = -0.817,\ r_4 = -0.778,\ r_5 = -0.258.$$

Thus it seems that our short-cut test fails to be decisive for this material. It might be added that according to the magneto-hydrodynamic theory of sunspots recently developed by H. Alfvén, the sunspot cycle is composed in two cycles, each of period 22 years.[4]

C. *Artificial data.* Finally, I shall report some results obtained when applying (1) to a series $x(t)$ which has been constructed according to the formula

$$x(t) = 5 \sin \frac{2\pi}{32} t + \xi_t \tag{4}$$

where the ξ_t are independent random variables, normally distributed ($m = 0$, $\sigma = 1$). It is seen that $x(t)$ is a series of rigid periodicity, with period 32. The ξ_t were taken from a table of random normal deviates constructed at the Institute of Statistics, Upsala.[5]

The series $x(t)$ constructed contains 14,000 items. Breaking up the series into 25 subseries of 560 items, each subseries contains 17 or 18 maxima. For each subseries, r_k was calculated for $k = 1, 2, 3, 4, 5$. Next Fisher's z-transformation was applied.

---

[4] See H. Wold and I. Galvenius, "Statistical Tests of H. Alfvén's Theory of Sunspots," *Arkiv Matem. Astron. Fys.*, 34 A (1947).

[5] H. Wold, "Random normal deviates," *Tracts for Computers,* XXV.

$$z_k = \tfrac{1}{2} \log (1 + r_k)/(1 - r_k).$$

For each k it turned out that the resulting 25 z_k-values were normally distributed, with fairly good approximation. The sample average and standard deviation of z_k are shown below. It is seen that the average values of z_k lie close to -0.549, which is the z which corresponds with $r_k = -1/2$. The observed standard deviations of z_k have been graduated by a linear expression in k.

	r_1	r_2	r_3	r_4	r_5
z_k: average	−0.521	−0.480	−0.512	−0.498	−0.534
z_k: stand. dev., obs.	0.244	0.198	0.290	0.335	0.326
z_k: stand, dev., grad.	0.22	0.25	0.28	0.31	0.34
r_k: Upper 5% level	−0.18	−0.14	−0.08	−0.04	+0.01
r_k: Upper 1% level	−0.04	+0.03	+0.10	+0.17	+0.24

Next, taking $z = 0.549$ for expected value of z_k, and making use of the graduated variances of z_k, Fisher's transformation was used to calculate 5% and 1% significance levels for r_k.

It is seen that the critical levels fail to separate the two hypothetical r-values, $-1/2$ and 0. For the artificial series (4), in other words, we must have for our data material covering much more than 17 periods to be able to decide between the two hypotheses by the use of the short-cut test.[6] To some extent the failure is due to the method, for since formula (1) makes use of but one observation in each period, the test is not sufficient in the sense of R. A. Fisher. To some extent, however, the difficulty lies in the problem—much material is always required if we are to decide between rigid or disturbed periodicity. With a series like Wolfer's sunspot numbers, for instance, I think it is doubtful whether any purely statistical test is able to bring out clearly, on the basis of the 17 cycles thus far observed, whether the periodicity should be regarded as rigid or disturbed.

---

[6] Judging from similar calculations, 35 periods would give satisfactory confidence levels.

Résumé

Le problème consiste à considérer des séries chronologiques présentant des oscillations ondulatoires, et à déterminer si les ondulations ont une périodicité rigide ou perturbée. L'analyse de la périodicité peutêtre faite avec des périodogrammes et des corrélogrammes; dans cette communication, toutefois, l'auteur présente une methode plus simple.

La méthode consiste à calculer le coefficient de corrélation r_k, comme il est défini par l'équation (1), représentant la corrélation entre la longuer des intervalles entre les maxima successifs. Dans certaines hypothèses, l'expérance mathématique de r_k est $-\frac{1}{2}$ si la série donnée est strictement périodique et 0 si la périodicité est perturbée. Le problème est donc transformé en un problème consistant à tester deux hypothèses alternatives concernant un coefficient de corrélation.

STATISTICAL METHODOLOGY SECTION

*Wednesday, September 10, at 2:00 p.m.**

CHAIRMAN :

J. R. Wormsley

Contributed Papers Bearing on the Development in the United States of Statistical Methodology in Engineering Research and Industrial Quality Control.

THE ROLE OF A STATISTICAL CONSULTANT IN A RESEARCH ORGANIZATION

Churchill Eisenhart
National Bureau of Standards (United States)

1. THE FUNCTION OF A STATISTICAL CONSULTANT

The function of a statistical consultant in a research organization is to furnish advice and guidance in the collection and use of data to provide quantitative foundations for decisions. For, as W. Edwards Deming has said, "...the chief duty of a statistician is to interpret data in such a way that they convey knowledge for purposes of prediction; another is to collect data in such a way that they provide maximum information; and still a third function is to help bring about some changes in the sources of the data. There is no use in taking data if you do not intend to do something about the sources of the data." Some may object that this attitude appears to be excessively pragmatistic—that study of the world to obtain knowledge for its own sake is an admissible occupation. True! But, it should be noted that the foregoing does not exclude description of the *status quo* as a means of understanding it, that is, to provide a basis for predicting the outcome of any course of action that could be taken. And research organizations generally take this attitude—they tend to agree with W. K. Clifford that "Every scientific fact is a shorthand expression for a vast number of practical directions: if you want so-and-so, do so-and-so." Research organizations in the main study the world for a dual purpose: (a) to determine how it can be changed; and (b) to decide which one of the possible worlds should be made the real one.

---

* Meeting arranged by the Washington Section of the American Society for Quality Control.

The tasks of a statistical consultant as a member of a research team are:

(1) *To furnish advice and guidance in the COLLECTION OF DATA.* This involves reaching agreement with the other members of the team on (a) *what* kinds of data are needed, (b) *how* they are to be collected, and (c) in *what quantity.* When (a) has been settled in the light of the objectives of the investigation, the knowledge of the workers in the substantive field concerned, and whatever supplementary information is accessible, it becomes the lot of the statistical consultant to adapt or develop, alone or in cooperation with other members of the research team, (i) patterns for the collection of the data (designs for experiments, patterns for the stratification of sampling, etc.) and (ii) methods of collecting the data (taking measurements, drawing samples, etc.) such that the effects of uncontrolled factors may be expected to balance out with increased replication, valid adjustments can be made for known or likely sources of bias, and the reliability of predictions based on the data obtained can be adequately estimated from the data alone or in conjunction with supplementary information.

(2) *To furnish advice, guidance, and (on occasion) assistance in the PROCESSING OF DATA prerequisite to their interpretation and presentation in summary form.* This requires knowledge of, or better, experience and skill with the procedures of statistical analysis and the tools of numerical mathematics. The former include the procedures appropriate to analysis of variance and covariance, contingency tables and goodness of fit, correlation (interclass and intraclass; simple, multiple, and partial; normal, serial, and rank), estimation by a single value (method of moments, maximum likelihood, "close") and by intervals (confidence intervals and fiducial limits), fitting frequency curves and their integrals (logits, probits), graduation and smoothing, least squares and regression equations (simple, multiple, and partial; linear and curvilinear; orthogonal polynomials), oscillatory time series (Fourier series and periodogram analysis; autoregressive series and correlogram analysis), rejection of observations, tests of significance (single-, double-, multiple-, and sequential-sampling; parametric and nonparametric); and so forth. The tools of numerical mathematics include computing machines, tabulating machines, mathematical tables, nomograms, and other aids to computation that can make arithmetic less burdensome, if not yet a pleasure. Clearly no man can be thoroughly acquainted, experienced, and skilled with all of these procedures and tools. But

there are times in the life of a statistical consultant when he wishes that he were!

(3) *To furnish advice and guidance in the INTERPRETATION OF DATA, and the PRESENTATION OF DATA as supporting evidence.* This involves (a) knowledge of the data (especially, *how* they were collected) and the objectives of the investigation, (b) some experience with data in the substantive field concerned, and (c) use of the mathematical theory of probability to determine with what confidence predictions of a given type can be made from data collected in a specified way and analyzed in a prescribed manner. It also requires (d) knowledge and experience in the presentation of data as evidence in tabular or graphical form; by averages, measures of dispersion, regression equations, etc.; by the representation of nature by probability distributions involving parameters that can be estimated from the data; and so forth.

The special contributions of the statistical consultant are his use of the mathematical theory of probability to determine what conclusions can be drawn and with what confidence, and his skill and experience in the presentation of data as evidence. The statistical consultant, in turn, is guided by the rule first stated explicitly, to my knowledge, by Walter A. Shewhart: "Original data should be presented in a way that will preserve the evidence in the original data for all the predictions assumed to be useful." This implies that if the data are presented in condensed form then their summary should not tend to "cause human action significantly different from what this action would be if the original [data] had been taken as a basis for evidence"—at least, so far as the set of "all the predictions assumed to be useful" is concerned. However, this raises the question: what is the set of "all the predictions assumed to be useful"? The answer appears to be: "We cannot have facts without some theory."

2. MAKING USE OF THE STATISTICAL CONSULTANT

(1) *As a collaborator in planning the procedure to be followed in an experiment, test, or survey.*

The cardinal rule for making effective use of the statistical consultant is

Call in the statistical consultant at the planning stage of an investigation. This rule is appreciated in the United States more widely today than formerly. Its importance derives from the fact that the decisive step in every investigation is the collection of the basic data: *what* data are collected is determined by the objectives of the investigation, and in turn determines its scope; *how* the data are collected determines the character of the predictions that can be validly based upon them.

Too often in the past the statistician has been called in "to the rescue" after difficulties have been encountered in the interpretation of the data —difficulties that could have been foreseen and avoided by appropriate planning of the collection of the data. In such cases, it is often impossible to draw any worthwhile conclusions from the data, with confidence. Sometimes a few results can be salvaged, but at great cost and with some risk. Cases in point are conclusions reached by analysis of variance when there are disproportionate frequencies in the respective classes: the arithmetic is almost certain to be heavy and the conclusions reached may depend on assumptions that have to be made in selecting the method of analysis—assumptions that cannot be checked from the data.

It is when the statistical consultant is called in at the planning stage that he can be most helpful in the processing and interpretation of data.

(2) *As an informal coordinator of the techniques of research.*

The statistical consultant performs another function, of supplementary nature, as he pursues his various tasks in the performance of his primary functions: the statistical consultant becomes an informal coordinator of research, drawing to the attention of workers in one branch of the organization related research completed, in progress, or contemplated in another branch—research which, as Warren Weaver put it, "involves the same verbs but different nouns." The statistical consultant is one of the means whereby techniques that have proved useful in one branch of science can be carried over and find useful application in other branches of science.

3. "STUDENT", THE FATHER OF STATISTICAL CONSULTANTS

In an address before the Royal Statistical Society in December 1934, R. A. Fisher stated that he found himself "more and more strongly drawn" to the opinion that

> "... the essential effect of the general body of researches in mathematical statistics during the last fifteen years is fundamentally a reconstruction of logical rather than mathematical ideas, although the solution of mathematical problems has contributed essentially to this reconstruction."

In his Harvard Tercentenary address he offers the following as an explanation of the delay in reaching this state of affairs:

> "Exact solutions of all the more important and immediate problems were possible by analytic methods certainly within the capacity of the greater writers of the last 150 years. That their existence remained for so long unknown, can only, I believe, be explained by the absence of any steady conviction that inferences involving an element of uncertainty *deserve* anything better than rough and approximate discussion."

It is my opinion that what he mentions undoubtedly had a retarding effect but that the following explanation is closer to the mark: For 19th century *physical science* the Classical Theory of Errors and Method of Least Squares were good enough. It was the advent of, study of, and experimentation with, first *biological* and later *industrial* material, subject to great and diverse forms of variation, that brought the need for statistical theory and methods of greater scope and effectiveness. And finally, given the need, the required ideas and methods were not forthcoming until perspicacious men of statistical inclination and mathematical bent were given an opportunity to don overalls, or at least laboratory frocks, and work side-by-side with those who were grappling with the problems to be solved.

The man who first clearly conceived the role of a statistical consultant in a research organization was William Sealy Gosset, employed as a Brewer by Messrs. Guinness of Dublin, and endeared to statisticians throughout the world as "Student", the pseudonym under which he published his far reaching contributions. Of "Student," his distinguished disciple R. A. Fisher has said:

> "How did it come about that a man of 'Student's' interests and training should have made an advance of fundamental mathematical importance, the possibility of which had been overlooked by the very brilliant mathematicians who have studied the Theory of Errors? Of his personal qualifications the most obvious are a clear head, and the practice of forming independent judgments.... One immense advantage which 'Student' possessed was his concern with, and responsibility for, the practical interpretation of experimental data. If more mathematicians shared this advantage there can be no doubt that mathematical research would be more fruitfully directed than it often is."

Résumé

La fonction d'un consultant en statistique dans une organisation de recherche est d'être conseil et guide dans la manière d'obtenir des données numériques et dans la façon d'en faire usage comme base quantitative de décisions. Cela implique la nécessité de se mettre d'accord, avec les autres membres du groupe engagés dans la recherche, sur les trois points suivants:

(a) la nature des données requises;
(b) la façon d'obtenir ces données;
(c) la quantité de données.

Les contributions particulières que peut faire le statisticien consistent dans son savoir et son expérience dans la façon de recueillir d'interpréter et de présenter les données comme matériel d'évidence, et dans l'emploi du calcul des probabilités pour déterminer quelles conclusions on est à même d'en tirer et quel est le degré de confiance que l'on peut attribuer à ces conclusions. Il est avantageux de recourir au service du statisticien dès le début c.à.d., lors de l'établissement du plan de recherche. Souvent le statisticien remplit la fonction d'un coordonnateur de recherche dans ce sens que les méthodes ou techniques jugées utiles dans une branche de recherche, par son intermédiaire, sont transplantées dans une autre branche où elles se prêtent à des applications fructueuses.

Discussion

Mr. Belz asked what statistical instruction is given to engineers in United States universities. He described the introductory work which is given in Australia, where engineering students are taught the possibilities of using probability theory and the principles and methods of quality control and simple experimental design.

EXPERIMENTAL DESIGN PRINCIPLES IN INDUSTRIAL RESEARCH

by George W. Brown
Research Professor, Statistical Laboratory,
Iowa State College (United States)

INTRODUCTION

Research, in this paper, is not confined to professional research groups or sections, but is taken in its broadest possible sense. Thus, the steps taken in initiating quality control procedures in a given process constitute research. Similarly, we use *experimental design* in a broad manner to cover systematically planned investigations, even when the origin of the data may not be experimental, so long as there is a sampling design based on a classification of sources of statistical variability.

The basic principles of experimental design, as laid down by Fisher and his followers, have become firmly established in the fields of agricultural and biological research. Industrial research has, on the other hand, been slow to take full advantage of these principles, partly because industrial workers do not have the same incentives to publish that academic workers have, and partly, perhaps, because industrial research is often an incidental part of an industrial executive's major operating duties. In some industries, also, executives may be frightened by the word "experimental."

The purpose of this paper is to present a general picture of the field for experimental design in industrial research, without attempting to compare the relative merits of specific designs in specific applications. Several widely differing phases of industrial activity will be considered, one by one, as candidates for the application of experimental design methods.

GENERAL PRINCIPLES

Experimental design starts with the classification of possible sources of statistical variability in one or more attributes, and generally leads to estimates of effects on the attributes in question, to estimates of the error of estimate, and to reclassification or search for further sources of variability. The goal of the design is to provide maximum information at a reasonable cost. It is generally the case that efficient designs lead to fairly simple analysis and interpretation of data, as a result of balances or orthogonalities inherent in the designs. The design is deliberately chosen to exhibit certain effects, in spite of other unavoidable disturbances to which the data are subject, and to pool, or confound, certain effects whose separate estimation is not required. It is evident that these goals are not peculiar to agricultural research, and there is a growing body of experience in industrial applications to testify to that fact.

INITIATION OF SAMPLING INSPECTION PLANS

The impetus of the late war has led industry to wholesale adoption of sampling inspection plans in the form of standard quality control procedures. Some companies now apply quality control methods throughout a plant process, from receiving department to shipping department. The initial steps in adopting such plans almost always involve research investigations to determine the magnitudes of "within lot" and "between lot" variation, thus providing opportunity for the application of experimental design methods. Some applications have been made here by engineers and others, originally unversed in statistical methods. These men have been attracted by successes obtained with the relatively simple quality control procedures and have been led to investigate the more sophisticated statistical techniques which make possible earlier and more intelligent application of the routine methods. In fact the whole problem of whipping an industrial process into "control" is a problem of locating sources of variability, finding the important ones and then taking action. Systematic investigation may be much more efficient than a blind attempt to reduce all variability everywhere to negligible size. It is important to realize that experimental designs need not interfere with normal production, since good design principles may often be met by a sampling design, without adopting experimental methods throughout, although planned experiments should not always be ruled out. Wherever sampling inspection is to be undertaken as a routine, there is an opportunity to apply experimental design methods. Take material which is to be inspected in a receiving department, and suppose that it is classified by lots, and by cases within lots. It is clear that more efficient inspection procedures can be imposed with a knowledge of the expected variability within cases and by case within lots, and by lot. If the material within cases is relatively homogeneous one would expect to take advantage of that fact by not sampling excessively within any individual case. This offers a beautiful opportunity to apply efficient statistical techniques, particularly when a single vendor is to supply similar material over a period of time.

It might also be mentioned that procurement of raw materials may be aided by planned investigations. Take the case of a plant which processes an agricultural product to extract a valuable oil. The oil content of the raw material is highly variable, and will depend on genetic factors, where grown, etc. In this situation a relatively small amount of well-planned investigation may repay its initial cost many times over.

Process Improvement

Research into process improvement has so far furnished perhaps most of the industrial applications of design principles. Investigators have used Latin squares, involving *n* operators, *n* machines, and *n* different products; there have been textile applications with randomized blocks and Latin squares. Perhaps the analogy with agricultural experimentation is more obvious here, where the object is to maximize the yield of a process, or the quality of output, and where a number of factors may enter. Perhaps, also, the possible gains to be derived from investigation are more obvious to plant management. There is hardly an industry which is not concerned with process improvement.

It is interesting to note the possible effects of studying lot to lot variation and within lot variation on research in process improvement. Such studies often turn up the key to better quality. A wartime study of a certain explosive led to improved quality by uncovering excessive variation which could be due only to improper packaging. In this particular case it was found that varying storage times and conditions brought about heterogeneity in material which was originally relatively homogeneous.

Laboratory Research

In these days research laboratories are maintained by large corporations, often for prestige value as much as for the actual output of the research group. In any serious research group, however, there are usually many opportunities for the application of experimental design techniques.

Take first the refinement of measuring techniques. Anybody who has ever subjected precision measuring equipment to searching investigation has discovered that they are subject to all the variability one can think of, and more. Precision techniques are normally sensitive to operator differences, day to day fluctuation, differences in type of material measured, etc. In industrial laboratories, the precision equipment is rarely as good as is claimed for it, unless considerable effort is expended on testing. Such matters as number and spacing of readings, frequency and method of calibration, depend on intimate knowledge of the behavior of the equipment. It is this author's opinion that many users of so-called precision equipment are not sufficiently conscious of the vagaries of such equipment. It is customary to assume that the equipment errors are small relative to the variability in the material under test. This may be true only when the equipment is used in the optimum manner, or when suitably modified.

Other applications in laboratory research may be similar to the applications in process improvement. A laboratory attempts to create new products with certain desired characteristics, with certain factors available as variables. In work based on a theoretical background there may be foundation for adopting theoretical shapes for response curves of one kind or another. In such cases, additional premiums are available for good design, making possible more efficient partition of the degrees of freedom of the experiment.

Other laboratory research may be concerned with consumers' use and maintenance of the company's product. The advantages of good experimental design are obvious here. Take, for example, a company which markets food products. Such companies frequently engage in research to develop new recipes and improve old ones, in order to encourage use of their product. The sensitivity of cooking recipes to various treatment times and quantities of ingredients, and the variability of people's tastes and judgments is well known. The efficient experimenter in this situation makes full use of the experimental design methods.

OTHER INDUSTRIAL RESEARCH

Research activities not previously classified include marketing and advertising research, and miscellaneous research problems such as the evaluation of the effects of training programs.

Marketing and advertising research are primarily dependent on the techniques of survey sampling, which in turn are based on experimental design principles. In these fields several research groups are active in applying up-to-date techniques to survey investigations. These fields also offer some interesting experimental possibilities which are normally not applied because of the understandable reluctance to experiment with "good will."

The evaluation of alternative training programs offers a field which so far as this author knows, is not widely exploited at the present time. The present interest in training methods, in large companies of all kinds, seems to indicate that some of these companies must soon begin to improve their methods of evaluation. The obvious questions of how much training, how long, and the related problems of selection of trainees can be answered only through well planned experimental work, which need not be large scale experiments. Some of the work done by and for the armed services during the war were very closely related to this industrial problem.

CONCLUSION

The reader can no doubt supply other examples, or classifications of applications not mentioned here. The important point here is not the difference in applications, but the degree of sameness. Stripped of the special cloak going with the specific application, the same goals remain, to obtain the maximum useful information with minimum effort, on the basis of classifications of sources of variation. The principles remain the same, the basic tools of good designs are the same, whether the investigations involve soy beans or automobiles.

Résumé

L'auteur traite les applications possibles des méthodes de plans d'expériences (*experimental design*) dans le domaine de la recherche industrielle, dans un sens assez large. Il donne un aperçu général, sans discuter des applications pratiques spéciales.

Bien que les principes fondamentaux des plans d'expériences aient été solidement établis dans le domaine de la recherche agricole et biologique, ils ont été beaucoup moins admis dans les applications industrielles. En principe, toutefois, les conditions favorables à l'application de ces théories ne sont pas limitées à la recherche agricole ou biologique.

L'introduction d'un programme de contrôle des fabrications est considérée comme une excellente occasion d'appliquer les méthodes des plans d'expériences, en procédant à des investigations systématiques des sources de variation, afin de raccourcir le temps nécessaire à mettre les procédés de fabrication sous contrôle. Dans ce cas, le plan d'expériences ne sous-entend pas des méthodes expérimentales, car les conditions de la théorie peuvent souvent être satisfaites par une méthode d'échantillonnage appropriée de la production courante.

La recherche en matière d'amélioration de procédés de fabrication a déjà donné lieu à de nombreuses applications, peut-être parce que l'analogie avec l'expérimentation agricole est plus apparente ici. Dans la recherche de laboratoire, l'amélioration de techniques de mesure présente, de même que les recherches concernant l'utilisation des produits par les consommateurs, une bonne occasion d'application des méthodes de plans d'expériences.

Parmi d'autres applications mentionnées par l'auteur, une possibilité intéressante semble être l'évaluation de programmes d'entraînement de personnel ainsi que le problème connexe de la sélection des candidats pour l'entraînement.

STATISTICAL QUALITY CONTROL IN AMERICAN INDUSTRY

by William R. Pabst, Jr.
Quality Control Division,
Bureau of Ordnance Department of the Navy
(United States)

Statistical quality control[1] is beginning to make a unique contribution to American industry. Over a wide area it is bringing the logic of modern statistics to bear on the day-to-day problems of the factory and assembly line. The testing of product qualities during production constitutes a type of continuing experiment in which sources of variation can be identified and eliminated if necessary. The Shewhart Control Charts, whether used on a single machine, a manual operation, or on a complicated assembly, embody in simplified pictorial form statistical tests of significance. Simple as these charts are, their use involves fundamental changes in concepts in the relation of engineering, production, and inspection.

Control charts are of two main types; first, charts for variables, known as "$\bar{x}$ and R charts"; and second, charts for attribute qualities, known as "p charts" for fraction defective and "c charts" for defects per unit. In $\bar{x}$ and R charts, the averages of samples taken usually at fixed intervals of time and the corresponding ranges or standard deviations are plotted. From accumulated data on ranges from successive samples, control limits based on the standard error of the mean are assigned about the averages. The control limits provide a graphic test, analogous to the analysis-of variance,[2] of the sub-lot to sub-lot variance in comparison with the within sub-lot variance for the past data; these limits are projected into the future as a means of detecting any changes in population from which the successive samples are drawn. Control limits for ranges or standard deviations provide a graphic counterpart of Bartlett's test for past data. In p and c charts, in which estimates of the mean and standard deviation are related, only sample percentage or number of defects is plotted and control limits are derived from accumulated averages.

---

[1] See W. A. Shewhart, *Economic Control of Quality of Manufactured Product,* D. Van Nostrand Company, New York, 1931. Also
Eugene L. Grant, *Statistical Quality Control,* First Edition, McGraw-Hill Book Company, Inc., New York and London, 1946.

[2] K. A. Brownlee, *Industrial Experimentation,* Directorate of Royal Ordnance Factories, Ministry of Supply, London, His Majesty's Stationery Office, 1946; and Henry Scheffé, "The Relation of Control Charts to Analysis of Variance and Chi-Square Tests," *Journal of the American Statistical Association,* Volume 42 (1947), p. 425.

Control charts may serve broader purposes of comparison when a proper selection of rational sub-groups is made. Control charts on each of two or more similar machines on the same operation provide a simple model for assigning variability as between operator and machine. Control charts also make possible a comparison of the intrinsic variability of machines or process with the specification tolerances and also a determination of the readiness of stand-by equipment. Modified control limits, derived from the subtraction of the appropriate standard error multiple from tolerance limits, furnish a continuing basis for the comparison between process variability and specification tolerances provided the underlying distribution is known.

Like all statistical tests, control charts may be used for a variety of purposes. They may be used, in the form of $\bar{x}$ and R charts, for control of a particular machine or group of machines. Percent defective and defects-per-unit charts may be used to evaluate the results of a complicated assembly operation such as building a radio chassis or filling a mail order package. Control charts may also be used to evaluate incoming product whether raw material or complex final product or sub-assembly. The closer statistical evaluation is to the process, the more rapidly can corrective action be taken when found necessary. Machine control charts make possible almost immediate action to correct process, whereas control charts used for evaluation of purchased product, as in acceptance inspection, may have delayed effect on processes exercised only through lot rejection or contract cancellation. In all these cases, the purpose of the statistical quality control chart is the test of the null hypothesis that the characteristics of the product or process under observation have not changed nor departed from satisfactory limits.

Well-developed statistical quality control installations include at least three features of quality evaluation; namely, evaluation of incoming material by means of acceptance sampling plans, evaluation of machine process or assembly by means of control charts, and evaluation of the quality of the outgoing product by means of sampling plans and attribute control charts. Control at one point on a process tends to show the desirability for control at many others. In an integrated company, control over the quality of the outgoing product of a given plant is usually of first consideration. Effective control of the final product imposes process controls at the intermediate stages of the process. Process controls imposed at strategic points tend to require additional controls on previous operations. Finally, control of incoming materials is required in order to ensure satisfactory quality level for further processing as well as for reasons of economy.

Acceptance sampling plans used both for acceptance of incoming material and for check inspection of outgoing product are based upon a classification of defects for the item and upon a structure of sampling plans. Quality control generally requires a more exact definition of product than is furnished by drawings and specifications. No matter at what place the production or process is to be evaluated, the primary step is the specification of what is to be measured, gauged, or judged and the weight that is to be applied in summing the quality of the end product. This is customarily given in the classification of defects for end product which is based on tolerances, finishes, and so forth for each part. The classification of defects translates the drawings and specification into a form upon which inspection procedures can be based. Each possible departure from drawing requirements is classed according to importance as judged by design engineers and as verified by laboratory test and production experience. Generally a fourfold classification of defects into Critical, Major, Minor A, and Minor B categories is used, and this is typically available to the suppliers of material. Sampling plans in widest use are the Dodge-Romig tables[3] and the Navy Tables.[4] Each of these provides a structure of sample sizes and acceptance and rejection numbers for given lot sizes and for various sampling risks. The Dodge-Romig Tables are indexed by Average Outgoing Quality Level (AOQL) and by Lot Tolerance Percent Defective (LTPD), whereas the Navy Tables are indexed by Acceptable Quality Level (AQL). The practice is becoming more common of including lot acceptability standards either in specifications or in purchased contracts. Indeed, some contractors have refused to accept orders from major companies until the lot acceptability levels and the classification of defects were established.

Such standardized acceptance inspection, based on quality control principles, has ramifications for both government and industry. Standard inspection tends to create a uniform evaluation of material making the anchor stamp, laboratory seal of approval, and similar designations capable of interpretation by those to whom material flows. In the future these stamps may be accompanied by the quality level at which material has been inspected. When this has been accomplished duplicate inspections at each handling of product may be avoided and the labor of re-inspection and wastage of materials can be eliminated. Standard inspection also makes it possible for those who purchase part of their requirement from others and make part of it in their own shops to have nearly identical inspection of own and purchased product at the same quality level. This implies, too, that government evaluation inspection is a necessary function even in its own plants and is not solely for the purpose of policing private producers.

---

[3] Harold F. Dodge and Harry G. Romig, *Sampling Inspection Tables*, John Wiley & Sons, Inc., New York, 1944.

[4] *General Specifications for Inspection of Material*, Standard Sampling Inspection Tables for Inspection by Attributes, U. S. Government Printing Office, Washington, D. C., April 1, 1946.

Machine control chart usage varies more than that of acceptance sampling plans. Some companies aim to establish a "chart on each machine." Others use charts only at points of strategic importance or when trouble is apparent. However, the tendency is to utilize charts at those points where critical or major defects (as shown by the classification of defects for the product) are affected, especially when the product is made in a company owned plant. Such controls on strategic points of process can also be suggested to suppliers. Some companies claim to have found it profitable to instruct their suppliers in statistical quality control methods. When machine control charts are used, exact specification of measurement and allowable tolerance in each operation must be provided. This, as well as the classification of defects, is the work of the design engineer. Design expectation can be checked against production experience by comparing machine variability and setting with tolerance limits. The installation of machine control charts on the manufacture of an item usually involves reconsideration of many of the specification details. This careful review of the specifications generally results in a relaxation of many more tolerances than need to be tightened to meet established performance requirements.

Actual practice in the construction and maintenance of control charts varies widely. Some companies have established the general practice of using a sample of five for variables charts and relating frequency of inspected sample to production levels. Others use various sample sizes. In some cases the American War Standards[5] are followed closely. However, most companies use modified control limits for process charts and some use chart control limits and modified limits together.[6] Modified limits (in the British sense) are found to be useful in shop work and in case of intermittent production. In these cases chart forms with precalculated modified limits are customarily sent to the operation with shop drawings and other instructions. The charts provide a means by which the quality control inspector can start immediately with the plotting of points in checking of product against specified requirements.

---

[5] *Control Chart Method of Controlling Quality during Production*, American War Standard ZI.3-1942, American Standards Association, 70 East 45th Street, New York, 1942; and *Guide for Quality Control and Control Chart Method of Analyzing Data*, American War Standard ZI.1-1941 & ZI.2-1941, American Standards Association, 70 East 45th Street, New York, 1941.

[6] E. H. Sealy, *A First Guide to Quality Control for Engineers*, Ministry of Supply, London, 1943.

In many applications of machine control charts, the procedure has generally been to secure control over one operation of importance. For example, in axle manufacture, a first attempt was made to eliminate the variation in shim required between the drive shaft and the carrier housing. To do this closer control of the mating pieces was needed and control over these was secured only by extending charts to approximately one hundred other dimensions. Each of these involved some changes in shop practices and in specifications.[7] The resulting product, made at a great saving in inspection time and rework, came from the cooperation of the engineering production, and quality control departments. Other applications also illustrate the tendency of controls at one point to reverberate throughout the system. Once the concept of statistical quality control has been thoroughly implanted for high levels in an organization, the need for quality evaluation throughout the entire process becomes recognized and the chain of control through the plant becomes extended.

Fundamental changes in design engineering, production, and inspection practices are usually found to be desirable with the installation of a comprehensive statistical quality control program. Many of the changes originate in the systematic organization and interpretation of information about the production processes and the checking of these against design requirements. Just as in the laboratory where the statistical adviser raises questions of technical people, the answers to which help the technical people solve their own problems, so the quality control engineer in requiring a systematic flow of information helps the technical staff coordinate its functions more fully.

The point has been made that the design engineer has the responsibility for preparing classifications of defects and for setting standards and tolerances for each dimension of piece parts required in the functioning or use of the device. Statistical quality control makes it possible for the production and inspection people to tell whether the process and machines specified can hold the design tolerances at economic cost and whether the sequence of such tolerances is proper. Many practical quality control men speak of having changes in tolerances on first operations affect vitally tool wear and possible tolerances on later operations. As the chain of control on a process is established, the necessary working limits on cutting and grinding in order to achieve necessary final surface dimensions can be worked out from experience. Conventional tolerances on many operations are found to be unworkable, uneconomic, or uncoordinated with other design requirements. All this brings the design engineer into more intimate contact with the operations in the shop, and with the verification of those operations in inspection. The crux of the system is to make the design engineer responsible for the complete specification of product, including the classification of defects and acceptable quality level, and it is generally not easy to make the design engineer accept this responsibility.

---

[7] "Statistical Quality Control," *Timken Axle News*, Volume 15, No. 2, 1947, published by Timken-Detroit Axle Company, Detroit, Mich.

Upon specification of the product, the inspection department undertakes the responsibility to measure the degree to which success has been accomplished. In this respect the inspection department becomes the agent of the design engineer in ascertaining that the design has been successfully achieved in production. Inspection thus is allied with the evaluation or production process and not with the improvement of production through screening inspection as has been too often the conception of inspection in the past. Screening inspection, when necessary, is required largely as a production function for bringing production up to necessary quality levels. Often screening inspection is checked for accuracy (based on the specification of the work unit of the screening operation involved) by means of sampling inspection. This changing attitude toward the inspection function as the evaluator of quality rather than the improver of quality is probably one of the most important consequences of the adaptation of statistical quality control. This concept has ramifications with respect to government inspection since it is becoming recognized that evaluation inspection by the consumer, whether it be government or a private company, has its purpose in comparison of product with standards set. Government inspection and adequate private inspection serve the same purposes.

Production functions in the plant are also reformulated once the specifications of successful production are set knowledgeably by design and enforced by inspection. Information derived from the evaluation of production is of vital interest in directing production process toward the successful accomplishment of the quality goals. Production accomplishments are related in meeting not only quantity standards but also quality specifications. That this is new is typified by the scrap-piles subject to disposal by plant salvage committees in the past, and the fact that production was allowed to continue in the absense of carefully specified goals and adequate means for evaluation of the process. Quality control in operation generally has made it possible to cut down the amount of scrap to very low limits and in some plants to entirely remove salvage departments. The reduction in scrap, although a tangible achievement, should not be considered as a major goal of quality control but as a minor derived benefit from better coordination of plant functions.

Coordination by management becomes easy when the primary plant functions are shown to be intimately related to production evaluation. The testimonial of each installation of quality control attests to this coordination. The following remark is typical: "A direct result of the introduction of the new methods has been the closer coordination of the function of the Inspection Department and the Production Department. Definite improvement in relationships is evident and inspection work is no longer thought of as a necessary evil but considered to be an inherent part of the manufacturing process."[8]

Many other changes in plant operation and plant practices may be found to be necessary after application of statistical quality control. Shop practices, such as the age old custom of producing to the basic dimension where a unilateral tolerance is specified becomes subject to reinterpretation under quality control. The unilateral tolerance has generally been related to making pieces of the "big" side since they could be cut down if necessary without wastage of the piece part. This however, is a holdover from the days of selective assembly. With adetequaach minery and with the help of quality control, it is found better to make the dimension within necessary tolerances as the first processing of the piece. The general policy found satisfactory under quality control is to use the middle of the tolerence range as the aiming point.

Additional gains from coordinated systems with statistical quality control are derived from dependence upon sample average rather than individual piece. When a machine is just capable of meeting tolerances, and individual measurement of pieces taken from a machine are relied on, chance high or low pieces tend to cause constant readjustment of the machinery. Frequent adjustment tends to cause the appearance of scrap whereas appropriate evaluation of the process can aid production of satisfactory material without down time for machine adjustment. One commercial company is reported to have extended the average time of machine adjustment from 20 minutes to 6 hours by this disregard of the individual piece and to have reduced its scrap by a large percentage. This disregard of the individual piece in favor of the average of a sample is probably the most difficult fundamental of statistics for the practical man to master.

These gains from clearer recognition of mutually interdependent functions within a producing company or organization give statistical quality control its greatest impetus. But they also tend to impose restrictions on its widespread and systematic application. Since it influences every activity in an organization, it cannot easily emerge from low levels in one of the existing departments but must have the full support of management. Piecemeal applications, despite the tendency of control to spread to prior operations, run into the stone wall or tacit opposition on the part of those who have not fully comprehended the system. For this reason many people at low levels in company organization have found it difficult or impracticable to apply the training derived from quality control courses. In many cases partial and inappropriate applications of control resulting in failure have been discarded after trial. On the other hand, even when management has accepted the principle of using statistical quality control, applications have not always done justice to the potential of the system. Many such systems simply provide for a review of screening inspection data without real benefit of statistical methodology. The elusive term "quality control" has so many possible facets of meaning that many have been led to feel that if they were paying any attention to factors affecting quality of product, as in painting machinery and sweeping benches, they were utilizing this effective new tool of control. Statistical evaluation may show many of these auxiliary aids to quality to be necessary, but they in themselves fall far short of an application of quality control.

---

[8] "Report of Statistical Quality Control at Naval Torpedo station, Alexandria, Virginia," October 7, 1946 (unpublished).

Sales departments have not been slow in organizing advertising campaigns around the words "quality control" whether or not any application of the statistical techniques of quality control are utilized. Moreover, gauge companies have not been free from blame for they have tended to associate quality control with the use of their gauges. This makes it extremely difficult to find out who is and who is not actually using statistical quality control in practice. It also points out the dilemma of the American Society for Quality Control as to the direction of its efforts, either toward selling management on general principles or on development of universally accepted procedures and the approval of quality control applications. Its attention has been directed largely to the first of these.

Applications of statistical quality control are now confined in large measure to the industrial leaders and to Army and Navy plants. Western Electric and Bell Laboratories, Johns Manville, Johnson and Johnson, Deere and Company, Minneapolis Honeywell Regulator, Timken Detroit Axle, Timken Bearing, Hoover Ball and Bearing, General Electric, Westinghouse, and Sentinel Radio, to mention a few, have developed more or less complete systems or plans for systems of statistical quality control throughout their plants. Many other companies appear to be engaged in extensive development and training programs. In this group, Ford, International Harvester, Sheaffer Pen, and Jack and Heintz, might be included. Details of some of these reach publications of the American Society for Quality Control,[9] or are discussed in the local section meetings.[10] An index to the extent of use of quality control is given in a recent tabulation of speakers available to local sections on the subject of quality control; the list included approximately 250 persons representing 100 industrial organizations.[11] This extends the list of those who participated in the wartime 10-day courses in quality control sponsored by the War Production Board. Nearly 2000 persons representing over 500 companies attended these courses and many of these have reported successful applications.[12] Nevertheless statistical quality control has not been applied in a great mass of American companies. In industrial cities few companies other than industrial leaders are utilizing statistical methods for controlling quality on a consistent basis.

Statistical quality control has survived the transition from its wartime use and is becoming a fundamental factor in peacetime production. It is expected that the example of the industrial leaders will be followed within the next decade by a large share of American companies. This will come partly from the pressure to reduce costs by elimination of rework and scrap, partly because the rapidly expanding use of acceptance sampling plans in setting standards will force the development of economical quality controls on suppliers of materials. Also the leadership of the Army and Navy and other governmental agencies in applying quality controls and setting quality standards may have a powerful influence. The primary factor restricting the use of statistical quality control is lack of knowledge of its effective use. This is being overcome in part by the increasing number of courses in statistical quality control and in statistical theory offered in technical schools, the development of training courses in these methods by large companies on a company wide basis, and the enthusiastic membership of the American Society for Quality Control. Managements of large industrial companies have enthusiastically responded to these new methods, but long and patient educational programs will be necessary to overcome the reluctance of older engineers, inspectors, master mechanics, and the foremen in the shops and to have them voluntarily embrace the fundamental changes in concepts that are necessary. What is important to the statistician is that the expansion of statistical quality control will introduce some features of industrial statistics to a new and large group of users, and will show the need for application of more precise statistical method to industrial and engineering problems.

---

[9] *Industrial Quality Control*, Vol. 1 through 4, first published July 1944.

[10] The American Society for Quality Control was organized on February 16, 1946, by accredited representatives from 17 local quality control societies. At the end of 1947, it was composed of 27 local Society Sections and included membership in excess of 2100.

[11] Mimeographed list distributed to local sections of the ASQC.

[12] Holbrook Working and Edwin G. Olds, "Organizations Concerned with Statistical Quality Control," *Quality Control Report* No. 2, June 1945; and Bibliography on Statistical Quality Control," *Quality Control Report* No. 11, Sep. 1945. Issued by Quality Control Program, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania.

Résumé

Le contrôle statistique des fabrications en série introduit dans l'industrie américaine des notions qui sont à la base des tests statistiques de signification, notamment celles comprenant les tests de l'homogénéité des moyennes et des variances des produits à certains intervalles pendant la production. Le contrôle statistique des fabrications en série tend non seulement à l'usage de cartes de contrôle de machines, à des méthodes de contrôle à la réception et de plans adéquats de spécification, mais aussi de la synthèse de ces éléments.

Les avantages de l'emploi de ces techniques ont en partie un caractère statistique en fournissant un plus grand degré de certitude pour des frais d'inspection donnés, ou en permettant de meilleures estimations des paramètres de la population par des méthodes rationnelles d'échantillonnage. En plus grande partie les avantages sont une conséquence des changements fondamentaux, nécessités par ces techniques, dans l'établissement des plans, des systèmes de production et d'inspection. Ces méthodes statistiques d'apprécier les produits et les procédés de production présupposent des définitions explicites de ces produits et procédés, ce qui a pour résultat une planification plus claire et la coordination des plans, de la production et de l'inspection. Le fait que les avantages proviennent dans une si large mesure des sources nonstatistiques, fait la force du contrôle statistique et en même temps, à un certain degré, la faiblesse de cette méthode. Son application est difficile sans l'appui complet de la direction principale de l'entreprise. Par conséquent, sa divulgation est limitée à cause du défaut de connaissance de ce nouvel instrument de contrôle de la part des directions et à cause de leur répugnance à apporter des modifications à des positions précairement équilibrées dans une organisation en fonctionnement.

Et, étant donné que les avantages initiaux sont attribuables en grande partie à une vision plus claire des devoirs et responsabilités mutuels, il peut arriver que les avantages possibles à long terme des applications systématiques du contrôle statistique soient négligés. Par conséquent, beaucoup de directions déclarent et s'imaginent que le contrôle statistique est appliqué dans leurs établissements, tandis qu'en réalité il ne s'agit que d'une version faible et impuissante des cartes de contrôle ou de l'inspection par échantillonnage qui est appliquée sans les changements d'organisation qui sont nécessaires à la réalisation de résultats effectifs. A présent, l'application effective du contrôle statistique des fabrications en série est, en général, limitée aux grandes entreprises industrielles, mais son influence est propagée par des cours d'entraînement dans des Universités et des Ecoles techniques, par les membres enthousiastes de l'"*American Society for Quality Control,*" par des articles instructifs dans des périodiques industriels et par l'impulsion donnée par le fait que les grandes entreprises et les services d'achat du gouvernement, pratiquent le contrôle à la réception. A la fin de la prochaine décade, la production industrielle non soumise au contrôle statistique sera peut-être aussi rare que l'expérimentation agricole sans planification statistique.

Discussion

Mr. Molina recalled an early application of mathematical theory through which shop practices of the Western Electric factories were almost completely revolutionized in 1905, when the design of automatic telephone systems was changed on the basis of probability and statistical theory.

Mr. Womersley asked to what extent modified control limits have become part of standard practice in the United States as they have in Great Britain, where they were used particularly during the war.

Mr. Pabst replied that the use of modified control limits, based upon the British practice, have been widely used in the United States. Usually, however, they are used jointly with regular control limits.

STANDARD SAMPLING-INSPECTION PROCEDURES

by W. Allen Wallis

University of Chicago (United States)[1]

1. Basic Concepts and Definitions

1.1 Nature of Sampling Inspection

Sampling inspection is a combination administrative technique and statistical technique. One of several possible actions is taken with respect to a set of distinguishable elements, choice of the action is based on measurements or classifications of one or more characteristics of individual elements, and although only a subset is measured or classified the action applies to the entire set. In these respects the problem of sampling inspection is essentially the fundamental problem of statistical inference [19].[2] But sampling inspection differs radically from other statistical techniques because the character of the set of elements with regard to which action is required may be profoundly influenced by the sampling procedure; in fact, its success in this administrative function is one of the principal criteria for evaluating a sampling-inspection procedure.

1.2 Terminology

An element is commonly an individual unit of a mass-produced commodity and is called an *item*—for example, a washer, a washing-machine, or a wash-cloth. The set then consists of a collection of such items, often having in common some such feature as having been produced under similar conditions or having been received in a single shipment, and is commonly called an *inspection-lot.* An item may, however, be such a thing as a single performance of a machine, for example, the quality delivered by a dispensing machine; the corresponding inspection-lot would be the aggregate of such quantities from repeated performances of the machine under stated conditions. Similarly, an item may be such a thing as the result of a clerical operation, for example, the product of a quantity and a price; the corresponding inspection-lot would be the aggregate of such products performed under stated conditions [1, 14, 30].

A subset that is measured or classified from an inspection-lot is called a *sample,* and the measuring or classifying is called *inspection.* The

---

[1] The author is indebted to Leonard J. Savage for a detailed and penetrating criticism of the manuscript that led to important improvements.

[2] Numbers in square brackets indicate references listed in Section 6 of this paper.

characteristics with respect to which an item is inspected are called *quality-characteristics*, and the measurements or classifications resulting from the inspection of an item are collectively called the *quality of the item* or *item-quality*.

Lot-quality is some specified function or group of functions of the qualities of the items in the inspection-lot. For example, it may be the arithmetic mean of one or more dimensions, the standard deviation of one or more dimensions, or the proportion meeting one or more criteria. *Sample-quality*, similarly, is some specified function or group of functions of the sample item-qualities.

Sampling inspection usually chooses between only two actions; for example, delivery of a lot may be accepted or it may be refused. Acceptance consists in taking one of the two actions, *rejection* in taking the other. It is conceived that knowledge of the lot-quality would determine which action should be taken.

In brief, *sampling inspection* is the process of accepting or rejecting material in lots on the basis of the quality of samples.

1.3 Acceptance Inspection and Process Control

The two most common uses of sampling inspection in industry are for process control and for acceptance inspection. The purpose of *process control* is to accept or reject a production process, that is, to decide whether items that will be produced in the future will be satisfactory. The purpose of *acceptance inspection* is to accept or reject existing material, that is, to decide whether items that have been produced in the past are satisfactory.

These two uses of sampling inspection have much in common, especially from the viewpoint of statistical theory, but there is an essential difference: process control is concerned with time. Thus a *control chart*, the primary tool of process control, involves samples similar to those of acceptance inspection, but for a single process it involves a series of such samples spread through time; and acceptance or rejection of the process at any moment is based on the recent temporal behavior of sample-quality as well as on its level at the moment. In acceptance inspection the temporal relations of items within a lot are ignored. As a result, acceptance inspection is in essential respects a different statistical tool than process control. If an inspection-lot contains only the output of a process that has been found by control chart analysis to be "in statistical control"[27], acceptance sampling serves no purpose (except perhaps as a sort of check, through more data, on the control sampling) [16]. On the other hand, acceptance inspection may be appropriate when a process is not in control or where process control is inappropriate or impossible—for example, when it is desirable or necessary to have "arms-length" inspection, that is, inspection entirely independent of the production process. It is often possible to improve the effectiveness of an acceptance inspection program by treating the sample-qualities of successive inspection-lots on a basis somewhat similar to process control.

The term "sampling inspection," like the term "quality control," is properly applicable both to acceptance inspection and to process control. In common usage, however, "sampling inspection" is apt to refer only to lot-by-lot acceptance inspection, and "quality control" only to process control. Perhaps a justification of this usage is that "inspection" suggests evaluation of existing results of past operations, while "control" suggests regulation of the potential results of future operations. At any rate, in this paper "sampling inspection" is ordinarily limited to "acceptance inspection" and does not include "process control."

Continuous process inspection has elements of both process control and acceptance inspection [2, 39]. It is particularly useful where lot-by-lot acceptance inspection is desired but accumulating the material

SAMPLING INSPECTION	THEORETICAL STATISTICS
inspection-lot	population (universe)
lot-quality	parameter
sample-quality	statistic
acceptance criterion (acceptable sample-quality)	critical region
acceptable lot-quality (acceptable quality level, *AQL*, or P_1)	parameter value specified by null hypothesis
rejectable lot-quality (lot tolerance percent defective, *LTPD*, or P_2)	parameter value specified by alternative hypothesis
producer's risk (probability that a lot of acceptable quality will yield a sample which fails to meet the acceptance criterion)	probability of an error of first kind (Type I error)
consumer's risk (probability that a lot of rejectable quality will yield a sample which meets the acceptance criterion)	probability of an error of second kind (Type II error)
operating-characteristic (probability, as a function of lot-quality, that an inspection-lot will yield a sample which meets the acceptance criterion)	power function, subtracted from 1

in inspection lots is impractical; such is the case with highly integrated, continuous processes, where raw material may remain in continuous motion through fabrication, packaging, and conveyor-belt loading into freight cars. Continuous process inspection will not be included, however, in the discussion of sampling inspection in this paper.

1.4 Sampling Inspection and Significance Tests

From the viewpoint of statistical theory, sampling inspection is a test of significance, that is, a test of the conformance of a set of data to a statistical hypothesis. The correspondence of terms is shown in the accompanying tabulation. The concepts denoted by these nine pairs of terms are fundamental to the current theory of testing statistical hypothesis. They were developed independently along closely parallel lines during approximately the decade 1925 to 1935, those of sampling inspection in the United States (almost entirely by H. F. Dodge, H. G. Romig, W. A. Shewhart, and their colleagues at the Bell Telephone Laboratories [3, 26]) and those of theoretical statistics in Great Britain (almost entirely by R. A. Fisher, J. Neyman, and E. S. Pearson, in large part at the University of London [7, 20]). Only in the past five to ten years, largely as a consequence of the war, have these two independent parallel developments tended to merge and to cross-fertilize.

2. LOT-QUALITY AND SAMPLE-QUALITY

2.1 Lot-Quality

In an overwhelming majority of cases where sampling inspection is used, lot-quality is defined in terms of the proportion of defective items in the inspection-lot, an item being defined as defective if any one of a list of attributes (*defects*) is present. But sometimes lot-quality is measured by the mean or variability of a quality-characteristic, for example, the average number of defects per item, the mean value of a continuously variable ("measurable") characteristic, or the standard deviation of a measurable characteristic.

Each method of measuring lot-quality permits a variety of useful methods of measuring sample-quality, provided certain information is known about the distribution of item-quality within inspection-lots. If the distribution of items by number of defects is Poissonian, for example, measurement of sample-quality in terms either of percent defective or of defects per item will permit inferences both about percent defective and about defects per item in the inspection-lot. Thus the method of measuring sample-quality is not in general fixed by the method of measuring lot-quality, but can be determined by economic considerations, in particular by comparing the statistical efficiency with the cost of each measure of sample-quality.

2.2 *Sample-Quality*

2.2.1 Binomial Distribution. Sample-quality is nearly always measured by proportion defective, each sample item being classified as defective (one or more defects) or nondefective. The binomial distribution then represents satisfactorily the sampling distribution of sample proportion defective, unless either the number of defectives or the number of nondefectives corresponding with the acceptance criterion constitutes a significant fraction of the number of defectives or nondefectives, respectively, in the lot, in which case the hypergeometric distribution must be used.

The binomial distribution has these advantages:

(i) It affords maximum operating simplicity and economy. Each item need only be classified into one of two categories; simple gauges and relatively little skill usually suffice, and there is a minimum of record keeping and computing.

(ii) It involves a minimum of mathematical assumptions. The binomial distribution (or the hypergeometric distribution under the conditions mentioned above) is certain to describe the distribution of sample-quality if (a) each item is simply classified defective or nondefective and (b) the sampling is effectively random.

(iii) Most of the computations necessary for devising and evaluating sampling plans are readily available, through tables of the binomial and incomplete-beta functions [23, 31, 32] through the normal [6] and Poisson [15, 41] approximations, which in turn are well tabulated, or graphically through binomial probability paper [18].

The principal disadvantage of the binomial distribution is statistical inefficiency; it requires larger samples than are needed when enough is known of the distribution to justify more detailed measurement of each item. For example, if a quality characteristic is normally distributed and values above a certain limit are defects, control of the percent defective through an appropriate function of the sample mean and sample standard deviation will not require as large samples as the same control through the sample percent defective [9, 29]. The larger samples required under the binomial, however, ordinarily cost less than would smaller samples involving more detailed measurement and more expensive calculation, chiefly because the cost per item is low with attributes but partly because with variables it is necessary to make careful statistical studies of the validity of the normality assumption (in fact, the greater the gain from using variables the stronger, generally speaking, is the dependence upon the normality assumption).

2.2.2 Poisson Distribution. If each item can have a large number of defects, and if the probability of its having any given defect is small and independent of whether it has other defects, then the Poisson distribution gives a satisfactory representation of the sampling distribution of number of defects [24]. Acceptance inspection is often based on number of defects for items of which many have several defects, for example, motor trucks, or telephone switchboards. If the mean number of defects per item is more than about 0.2, use of the Poisson distribution permits appreciably smaller samples than would be required for the same reliability with the binomial distribution [30]. On the other hand, special studies are required to determine whether the distribution of defects is in fact Poissonian, so inspection on this basis is not so readily standardized and routinized as inspection on a binomial basis.

2.2.3 Normal and Student Distributions. When lot-quality is measured by the mean of a characteristic, the sample mean of this characteristic is commonly used as the measure of sample-quality. The normal distribution gives a satisfactory representation of the sampling distribution of sample means, provided the lot standard deviation is known [25]. Sometimes the lot standard deviation is stable, and can therefore be determined from past data. More often, the standard deviation must be estimated from the same sample that is being used to test the mean, in which case the Student distribution is appropriate [8, 21]. In general, these distributions are not seriously affected by moderate departures from normality in the distribution of items, so sampling inspection is less dependent on continual verification of mathematical assumptions when done on this basis than when done on any other basis except the binomial.

2.2.4 Chi-square and Range Distributions. It is not often that lot-quality is measured by the standard deviation. When it is, sample-quality is usually measured either by the sample standard deviation or by the sample range. The chi-square distribution represents the sampling distribution of the sample standard deviation (more precisely, of a simple function of the sample standard deviation), if the items are normally distributed [8, 29, 40]. For very small samples the range may be used as the sample measure of variability; its distribution is also known, again assuming normality [12, 22]. Acceptance inspection, unlike process control, ordinarily involves samples too large for the range to be efficient, despite the simplicity of measuring it. The average range in subsamples is a device for reducing this inefficiency [11]; another device is an appropriately chosen inter-percentile range, for example, the difference between the first and ninth deciles [17].

2.2.5 Noncentral-t Distribution. If lot-quality is measured by the proportion of items for which a measurable quality characteristic exceeds (alternatively, falls short of) a stated limit, and if item-quality is normally distributed, then sample-quality may be measured by the difference between the limit and the sample mean, divided by the sample standard deviation. This measure of sample-quality will be distributed according to the noncentral-*t* distribution [9, 13, 29]. Substantial reductions in sample size can be obtained by basing inspection on this distribution instead of on the binomial distribution, but the greater inspection cost per item when measurements are required, the greater volume and complexity of records and computations, and the necessity of competent statistical checks on the normality assumption, all combine to limit its usefulness to products for which inspection costs per item are especially high even with attributes, as when the inspection involves destruction of expensive items.

3. TYPES OF SAMPLING

3.1 *Randomness*

Each of the four types of sampling—single, double, multiple, and sequential—commonly used in sampling inspection is based on randomness. That is, each individual sample must be drawn by an effectively random process. Randomness is an attribute of a sampling process, not of an individual sample; a sampling process need not be strictly random but only effectively random; and the samples consist of sets of measurements or classifications, not of the items to which the measurements or classifications pertain.

A sampling process is ***strictly random*** if the probability that any item-quality will be included in the sample is (i) equal to its relative frequency in the inspection-lot, and (ii) independent of the other item-qualities in the sample, except through their number. One way to assure strict randomness is to see that each item in the inspection-lot has the same chance of being selected for inspection, and that this chance is not affected by which other items are selected; this method assures strict randomness with respect to every characteristic of the items.

Even if a sampling process is not strictly random, it may be ***effectively random***, that is, it may be that the probability of drawing a sample with any given value of a specified measure of sample-quality is the same as it would be under strictly random sampling. For example, if an inspection-lot is presented with the items arranged in a number of piles, it may be that taking the top item from each pile would be an effectively random sampling procedure with respect to, say, mean tensile strength. Whether a particular sampling process is effectively random is itself a subject requiring statistical investigation, unless strict randomness has been obtained by assuring each item the same probability of representation in the sample.

Most discussions of randomness simply say that the probability of inclusion must be the same for all items in the population. Since this condition is often impractical, it is perhaps worth restating the point that, while sufficient, it is not necessary: Given the item-qualities of an inspection-lot, consider a hypothetical population each of whose elements is a combination of N of the item-qualities, each possible combination of N item-qualities being equally frequent in this population. Then consider each combination of N to be replaced by a single number, representing whatever measure of sample-quality is of interest. Thus, we have generated a population of sample-qualities; the relative frequency of any sample-quality in this population shows what the probability of that sample quality would be with strictly random sampling from the inspection-lot. Now a sampling process may be termed effectively random if the probability that it will yield a sample with any specified sample-quality is equal to the relative frequency of this sample-quality in the hypothetical population of sample-qualities. Expressed differently, a sampling process is effectively random if the population of sample-qualities that would be generated by infinitely many repetitions of the sampling process is identical with the population of sample-qualities derived from our hypothetical population of combinations of N items from the inspection-lot. Thus, a sampling process is effectively random if it produces the sample-qualities in the correct proportions, regardless of the internal compositions of the samples that accomplish this.

3.2 Single and Sequential Sampling

Two essentially different types of sampling are widely used in sampling inspection. In one type, ***single sampling***, a fixed number of items to be inspected is determined in advance and all of them are inspected; in the other type, ***sequential sampling***, the number of items is a variable, and depends upon the sequence of item-qualities found in sampling. Sequential sampling may take various forms; in particular, there may be an upper limit to the number of items that will be inspected or this may be open, and the successive samples may contain only one item each or they may contain several items. When not more than two samples are to be inspected, we have ***double sampling*** (historically the first systematic type of sequential sampling), and when not more than some fixed number (greater than two) of samples are to be inspected we have ***multiple sampling***; the individual samples in both double and multiple sampling almost invariably contain several items rather than only one. ***Curtailed sampling*** refers to discontinuing inspection within one of the samples, instead of inspecting an integral number of samples; inspection may be curtailed if the outcome of the sampling is determined before the entire sample has been inspected, as for example, when the number of defectives reaches the rejection number. Curtailed single sampling, if the preceding distinction between single and sequential sampling is adhered to, is sequential rather than single sampling, since the amount of inspection is a variable depending upon the progress of the sampling. The term "sequential sampling" is often used in a special sense to mean either (i) a particular sequential form called the "sequential probability ratio" form in which the acceptance and rejection criteria are essentially parallel linear functions of the sample size [30, 38], or (ii) multiple sampling, usually with a fixed maximum number of samples, in which the acceptance and rejection criteria are derived, at least to a first approximation, from the sequential probability ratio [28, 34, 35, 36, 37].

Single sampling can minimize (i) the maximum number of items inspected per inspection-lot, (ii) the number of separate samples inspected, and (iii) the variability of the number of items inspected per inspection lot. Sequential sampling can minimize (iv) the average number of items inspected per inspection-lot.

4. STATISTICAL PROPERTIES OF SAMPLING PLANS

4.1 *Operating Characteristics*

A sampling plan is essentially a rule for accepting or rejecting inspection-lots. In the preceding sections we have considered the main features of such a rule, in particular selection of samples, computation of the sample-quality, and acceptance and rejection criteria. Whatever the details of the rule, and however naive or sophisticated the motivation for them, a most important aspect of their statistical consequences is described by the operating-characteristic function, called the *OC* function or the *OC* curve.

The *OC* function shows the probability that the sampling plan will result in acceptance: the probability, in other words, that random sampling will result in a sample-quality meeting the acceptance criterion. Obviously this probability depends upon the lot-quality; with any useful sampling plan it is large for very good lots and small for very poor lots. A graph showing the probability that an inspection-lot will be accepted under a given sampling plan as a function of the lot-quality is the *OC* curve of that sampling plan.

The *OC* function shows a buyer or other receiver of the goods what protection is afforded by the sampling plan, that is, what the probability is that if a poor lot is submitted it will be accepted. It shows a seller or other supplier of the goods what is required of him by the sampling plan, that is, what his quality must be in order to achieve a given probability of having his material accepted. Given a sampling plan, the seller will strike a balance between two factors: (i) incurring the costs of producing high quality material, but thereby avoiding the costs of having numerous inspection-lots rejected; (ii) incurring the costs of numerous rejections but avoiding the costs of high quality production. The buyer, in selecting a sampling plan, will also strike a balance between two factors: (1) incurring substantial costs for unsatisfactory material that is passed by the sampling plan, but avoiding a high purchase price; (ii) incurring a high purchase price, but avoiding high costs from accepting unsatisfactory material. This conscious reaction of the supplier to the *OC* function, and the fact that the receiver is influenced in selecting an *OC* function by his estimate of the supplier's reaction, is a vital feature of sampling inspection, and makes it much more than a useful statistical technique: it is an effective administrative technique for coordinating two independent parties to an undertaking without the necessity of an authority empowered to issue commands to both.

It should be emphasized that the *OC* function does *not* show the probability that an accepted lot will have a certain quality. That probability would be proportional to the product of two probabilities, (i) the probability that a lot of the specified quality will be offered, and (ii) the conditional probability that if such a lot is offered it will be accepted. Only the second of these is subject to control by an acceptance inspection plan, and it alone is shown by the *OC* function. Any information about the first probability will profoundly influence a supplier's or a receiver's reaction to the operating characteristics of a sampling plan; indeed, the buyer's and seller's compromises between alternative costs, referred to in the preceding paragraph, cannot be made rationally without assumptions about the probabilities, with various sampling plans in effect, that various lot-qualities will be submitted.

While a sampling plan cannot control the outgoing quality of each individual lot, an ingenious modification often makes it possible to limit the worst possible *average* outgoing quality. This modification consists of a requirement that rejected lots be replaced by perfect lots, by inspecting every item in each rejected lot. Then outgoing quality will be perfect if incoming material is either all good or all bad, and there will be some intermediate quality which if invariably submitted, would minimize the average outgoing quality; the minimum that would be attained is called the *average outgoing quality limit* or *AOQL* [5]. Such a modification is not possible if inspection is destructive or if virtually perfect accuracy in inspection is not feasible; and, as remarked, the limit applies only to the average quality of a large number of lots, not to individual lots. Like continuous process inspection, inspection on an *AOQL* basis has elements of both acceptance inspection and process control.

4.2 Amount of Inspection

The cost of operating a sampling plan depends largely on the amount of inspection it requires. With sequential sampling the amount of inspection varies with the lot-quality of the submitted material, and for a given lot-quality it varies by chance. Thus, for any given lot-quality there is an expected or average amount of inspection required to reach a decision, and this average amount of inspection (usually called the average sample number, or *ASN*) is a function of lot-quality whose graph is called the *ASN* curve [28, 30]. For single sampling, the *ASN* function is, of course, constant, that is, independent of lot-quality.

The *ASN* function shows an important cost characteristic of a sampling plan. Other characteristics of sampling plans that affect their operating costs include [28, 29]:

(i) the maximum number of samples that may have to be inspected;
(ii) variability in the number of samples;
(iii) the maximum number of items that may have to be inspected;
(iv) variability in the number of items;
(v) the cost of the individual measurements;
(vi) the amount and complexity of record keeping and computing;
(vii) the cost of any statistical studies necessary to verify assumptions about the distribution of sample-quality, the effective randomness of sampling procedures, the validity of tests of increased severity, etc.;
(viii) the inspection-lot size, since for a given sample size the amount of inspection *per unit produced* is inversely proportional to the inspection-lot size.

An offset to the cost of a sampling inspection plan is the value of any byproduct information it yields—as, for example, if information on number of defects per unit, or on kind of defects is of value, even though acceptance or rejection could be based on sample percent defective alone.

When the *ASN* function and other pertinent data indicate that the operating costs of a sampling plan are too high, one or more of the following adjustments may be possible:

(i) A similar plan, but with less protection to buyer or seller may be adopted.
(ii) A plan may be adopted which uses more detailed quality information for each item, as when measurements and the non-central-t distribution are substituted for attributes and the binomial distribution.
(iii) A change may be made in the quality-characteristic being inspected, substituting a characteristic which is functionally related to, or at least highly correlated with, the characteristic of primary concern. The most common substitution is a test

of increased severity, since less inspection is required when the acceptable and rejectable lot-qualities are, for example, 5 and 15 percent defective than when they are 0.05 and 0.15 percent defective.

(iv) The plan may be kept unchanged but the inspection-lot size increased. The *OC* function, except when the sample is large relative to the inspection lot, depends almost entirely on the absolute size of the sample. Increasing the inspection-lot size therefore leaves the protection unchanged but reduces the amount of inspection per item produced.

5. STANDARD PROCEDURES

5.1 *General Principles*

The discussion so far has dealt with sampling plans—their nature, their consequences, and their costs—and has outlined the chief considerations that would be taken into account by a statistician carefully designing a sampling plan to suit the many facets of a particular situation. Such careful designing of sampling plans is often impossible because of scarcity of sampling inspection experts, and even when possible it would often be uneconomic.

A standard sampling inspection procedure is a procedure for selecting and operating sampling plans. With a standard procedure, many questions are settled arbitrarily which would be decided by an expert only after careful study; for other questions, the range of choice is narrowed to relatively few alternatives, with detailed prescriptions for choosing among these; and some questions are analyzed into component parts, each to be decided separately by someone whose regular work gives him the competence. As with mass production, the object of standardization is to increase output and lower cost; as with mass production, some loss of quality is incurred through standardization, but usually not nearly enough (if the standardizing is sound) to offset the gains in quantity and cost. In other words, a standard sampling inspection procedure makes it possible to introduce useful, but usually not ideal, sampling inspection plans rapidly and economically.

The foundation of a standard sampling inspection procedure is a set of sampling plans—sample sizes and acceptance and rejection criteria—together which such pertinent information about each plan as its acceptable quality level, rejectable quality level (often called lot tolerance percent defective or *LTPD*), operating-characteristic function, average outgoing quality limit, and average sample number function. Such a set should include a plan sufficiently close to the ideal one for any specific circumstances, but should contain as few different plans as will meet this requirement reasonably satisfactorily. Numerous questions must be decided in order to find in this set of sampling plans the one most appropriate to a particular problem. A standard procedure lists these questions, allocates the responsibility for answering them, and provides guidance for answering each question.

An important principle in constructing a standard procedure is that the component questions into which the selection of a plan is resolved should be as nearly as possible mutually orthogonal: the effect of a certain answer should be the same whatever answers are given to other questions. Orthogonality is necessary for the administrative workability of the procedure. Suppose, for example, that the probability of accepting material at the acceptable quality level were to depend on whether single or sequential sampling is used. Then the person responsible for determining the acceptable quality level could not properly do so unless he knew what type of sampling would be chosen. Choice of the acceptable quality level is appropriately made by a technical person familiar with the production and uses of the commodity, while choice of the type of sampling is appropriately made by an inspection person. It would be necessary for these two persons to try to strike a balance between noncomparable factors. With a properly designed standard procedure, however, the probability of accepting an inspection-lot of acceptable quality will be substantially independent of the type of sampling.

A standard sampling inspection procedure can be divided into four major steps [28, 36]:

(i) Establishment of standards.
(ii) Installation of a sampling plan.
(iii) Operation of the sampling plan.
(iv) Review of results.

The first and fourth steps relate primarily to quality requirements of the product while the second and third relate primarily to the administration of sampling inspection.

5.2 *Special topics*

5.2.1 Sample Size. The operating-characteristic function of a sampling plan, which fully describes the protection afforded by the plan, may be characterized by two features, its steepness and its location. The steepness indicates the degree of discrimination attained: if the *OC* function is steep, there will be only a small gap between qualities that are nearly always accepted and qualities that are nearly always rejected; if it is shallow this gap will be large. The location indicates which qualities are generally accepted and which are generally rejected.

Since the steepness of the *OC* function is determined almost entirely by the sample size, and scarcely at all by the lot size (except under conditions noted previously), it might seem that the same sample size would be used for all lot sizes, this sample size depending on the sharpness of discrimination required. How much discrimination it is wise to buy depends, however, on the cost; the point at which further discrimination would not be worth its cost comes sooner when the cost is high than when the cost is low. The cost that is relevant here is the cost of inspection per item produced. This will be less if inspection-lots are large than if they are small. Thus, more inspection in absolute amount, but presumably less in relative amount, should be given to large lots than to small lots, other things being equal [28]. The proper increase of sample size with lot size is one of the most difficult problems in designing special sampling plans; in current standard procedures it is simply determined arbitrarily, with no other guide than that there should be some increase but not as much, proportionately, as in the lot size [28, 33, 34, 35, 36].

5.2.2 Acceptable Quality Level. The fact that the sample size, and hence the steepness of the *OC* function, increases with lot size means that complete uniformity of treatment is not possible for all classes of suppliers. For example, if all are given the same probability of acceptance for good material, suppliers submitting small inspection-lots will have a higher probability of acceptance for poor material than will suppliers submitting large inspection-lots. On the other hand, if the receiver insists on uniform protection against the acceptance of poor material, suppliers submitting large inspection-lots will have a higher probability of acceptance for good material than will suppliers submitting small inspection-lots.

It is to the receiver's interest that inspection-lots be as large as practicable, since the larger they are the steeper the *OC* function and the less the inspection costs. The supplier's interest is in small inspection-lots if all inspection-lots have the same probability of acceptance for good material, and in large inspection-lots if all inspection-lots have the same probability of acceptance for poor material. If all suppliers were essentially similar in their capacity for submitting large inspection-lots, this would point toward holding the consumer's risk constant. In fact, suppliers generally are not similar in this regard, and holding the consumer's risk constant would discriminate against small suppliers and in favor of large suppliers. Current standard procedures, therefore, generally hold the producer's risk essentially constant; that is they treat alike (give practically the same probability of acceptance to) all inspection-lots that are of acceptable quality [28, 33, 34, 35, 36].

It is thus convenient in standard procedures to measure the location of an *OC* function by, and to classify sampling plans according to, some lot-quality which will be accepted nearly always. "Nearly always" requires exact definition; the usual definition in this context is "approximately ninety-five percent of the time." Thus sampling plans are characterized and classified by acceptable quality level (AQL) defined as that lot-quality for which the probability of acceptance is 0.95.

5.2.3 Type of Sampling. When different types of sampling are included in a standard procedure, for example, single, double, and multiple sampling, their *OC* functions are matched as nearly as possible. That is, each available *OC* function can be attained by single, by double, or by multiple sampling. The decision as to type of sampling can be made by inspection personnel and changed by them at will, since it will affect only the costs of operating the plan but not its consequences.

5.2.4 Review of Past Results. Standard procedures usually provide for analyzing the results of a series of inspection-lots on something resembling a process control basis. If evidence accumulates that bad lots are not likely to be submitted, as when the process appears to be, in some sense, "in control", the protection against such lots can be reduced (*reduced inspection*); that is, the *OC* function can be flattened, keeping the *AQL* fixed, which reduces the amount of inspection. Similarly, if evidence accumulates that bad lots are very likely to be submitted, the protection against accepting them can be tightened (*tightened inspection*); tightening may be accomplished either by increasing the amount of inspection with the *AQL* fixed, or by decreasing the *AQL* with the amount or inspection fixed, the latter being perhaps justifiable on the ground that the receiver should not incur extra inspection costs because of the supplier's shortcomings [28, 36].

5.3 Prospective Developments

Development of standard sampling inspection procedures in the immediate future is likely to proceed along four lines.

(i) *Revision in detail.* Most existing standard procedures have been prepared hurriedly. Approximations have been used in computing *OC* functions. Some sets of supposedly matching plans may not actually match as closely as possible. Plans may have been matched too closely for different types of sampling at the expense of consistency among acceptable quality levels and among lot sizes. No existing standard procedure supplies the *ASN* curves for its sampling plans. Criteria for shifting to reduced or tightened inspection, or for returning to normal inspection, are probably not ideal. In general, there are many opportunities for improving existing standard procedures.

(ii) *Integration with process control.* Developments in acceptance inspection and in process control have for the most part, proceeded more or less independently. Especially where acceptance inspection is used by a manufacturer within his own plant or among his own plants, there appears to be need of more effective integration of the two than has so far been worked out.

(iii) *Extension of coverage.* Existing standard procedures utilize only the binomial distribution, treat different quality-characteristics essentially independently, and require the accumulation at one time of the full inspection-lot. Techniques for removing all these and other limitations are now available and in use in special sampling plans, but have not been incorporated into standard procedures. With an increase in the number and quality of statisticians available, it may be possible to utilize routinized and standardized procedures for continuous process inspection, for the use of variables to control proportion defective, and for multivariate measurement of sample-quality [2, 9, 29, 30, 39].

(iv) *Development of the economic theory of sampling inspection.*
Even where a sampling inspection plan is designed for a particular situation, the fundamental problems of the economically best amount of inspection and location of the AQL cannot be solved with existing theory. Until more progress is made on these basic economic issues, standard procedures will remain fundamentally arbitrary in their most important aspects. The progress to be made in standard sampling inspection procedures from a solution of these economic issues probably dwarfs the combined potentialities of all three of the other lines of development [28].

References

[1] Deming, W. Edwards, and Geoffrey, Leon, "On Sample Inspection in the Processing of Census Returns," *Journal of the American Statistical Association*, 36 (1941), 351-60.

[2] Dodge, H. F., "A Sampling Inspection Plan for Continuous Production," *Annals of Mathematical Statistics*, 14 (1943), 264-79.

[3] Dodge, H. F., and Romig, H. G., "A Method of Sampling Inspection," *Bell System Technical Journal*, 8 (1929), 613-631.

[4] Dodge, Harold F., and Romig, Harry G., *Sampling Inspection Tables: Single and Double Sampling*, New York (John Wiley and Sons) 1944, pp. vi & 106.

[5] Dodge, H. F., and Romig H. G., "Single Sampling and Double Sampling Inspection Tables," *Bell System Technical Journal*, 20 (1941), 1-61.

[6] Feller, W., "On the Normal Approximation to the Binomial Distribution," *Annals of Mathematical Statistics*, 16 (1945), 319-29.

[7] Fisher, R. A., *Statistical Methods for Research Workers*, First and Later Editions, Edinburgh and London (Oliver and Boyd) 1925 and later.

[8] Freeman, H. A., *Industrial Statistics: Statistical Technique Applied to Problems in Industrial Research and Quality Control*, New York (John Wiley and Sons) 1942, pp. x & 178.

[9] Goode, Henry, *Sampling Inspection by Variables when Lot Quality is Characterized by Proportion Defective*, Stanford, California (Department of Mathematics, Stanford University) 1947, pp. ii & 37 & (xxxvii). Hectographed and lithoprinted.

[10] Grant, Eugene L., *Statistical Quality Control*, New York and London (McGraw-Hill Book Co.) 1946, pp. xii & 563.

[11] Grubbs, Frank E., and Weaver, Chalmers L., "The Best Unbiased Estimate of Popuation Standard Deviation Based on Group Ranges," *Journal of the American Statistical Association*, 42 (1947), 224-41.

[12] Gumbel, E. J., "The Distribution of the Range," *Annals of Mathematical Statistics*, 18 (1947), 384-412.

[13] Jennett, W. J., and Welch, B. L., "The Control of Proportion Defective as Judged by a Single Quality Characteristic Varying on a Continuous Scale," *Supplement to the Journal of the Royal Statistical Society*, 6 (1939), 80-8.

[14] Jones, Howard L., "Sampling Plans for Verifying Clerical Work," *Industrial Quality Control*, 3 (1947), 5-11.

[15] Molina, E. C., *Poisson's Exponential Binomial Limit*, New York (D. Van Nostrand Co.) 1943, pp. (viii) & 47.

[16] Mood, Alexander M., "On the Dependence of Sampling Inspection Plans Upon Population Distributions," *Annals of Mathematical Statistics*, 14 (1943), 415-25.

[17] Mosteller, Frederick, "On Some Useful 'Inefficient' Statistics," *Annals of Mathematical Statistics*, 17 (1946), 377-408.

[18] Mosteller, Frederick, and Tukey, John W., *Binomial Probability Paper*, Norwood, Mass. (Codex Book Co.) 1947.

[19] Neyman, Jerzy, "Inductive Reasoning or Inductive Behavior: Modern Concepts of Mathematical Statistics," these Proceedings (III,-International Statistical Institute), 423-431.

[20] Neyman, J., and Pearson, E. S., "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society of London*, A, 231 (1933), 289-337.

[21] Peach, Paul, *An Introduction to Industrial Statistics and Quality Control*, Second Edition, Raleigh, N. C. (Edwards and Broughton) 1947, pp. xvi & 236.

[22] Pearson, E. S., and Hartley, H. O., "The Probability Integral of the Range in Samples of n Observations from a Normal Population," *Biometrika*, 32 (1943), 301-8.

[23] Pearson, K. (editor), *Tables of the Incomplete Beta-Function*, London (Biometrika Office, University College) 1934, pp. lix & 494.

[24] Przyborowski, J., and Wilen'ski, H., "Statistical Principles of Routine Work in Testing Clover Seed for Dodder," *Biometrika*, 27 (1934), 273-92.

[25] Romig, Harry G., *Allowable Average in Sampling Inspection*, New York (Doctoral Dissertation, Columbia University) 1939, pp. 58.

[26] Shewhart, Walter A., *Economic Control of Quality of Manufactured Product*, New York (D. Van Nostrand Co.) 1931, pp. xiv & 501.

[27] Shewhart, Walter A., with the editorial assistance of Deming, W. Edwards, *Statistical Method from the Viewpoint of Quality Control*, Washington (Graduate School, Department of Agriculture) 1939, pp. x & 155.

[28] Statistical Research Group, Columbia University (edited by H. A. Freeman, Milton Friedman, Frederick Mosteller, and W. Allen Wallis), *Sampling Inspection: Principles, Procedures, and Tables for Single, Double, and Sequential Sampling in Acceptance Inspection and Quality Control Based on Percent Defective*, New York and London (McGraw-Hill Book Co.) 1948, pp. xx & 395.

[29] Statistical Research Group, Columbia University, (edited by Churchill Eisenhart, Millard W. Hastay, and W. Allen Wallis), *Selected Techniques of Statistical Analysis for Scientific and Industrial Research and Production and Management Engineering*, New York and London (McGraw-Hill Book Co.) 1947, pp. xiv & 473.

[30] Statistical Research Group, Columbia University (edited by H. A. Freeman, M. A. Girshick, and W. Allen Wallis), *Sequential Analysis of Statistical Data: Applications*, New York (Columbia University Press) 1945-6, pp. xxxviii & 367.

[31] Thompson, Catherine M., "Percentage Points of the Incomplete Beta Function," *Biometrika*, 32 (1941), 168-81.

[32] U. S. Army Security Agency, *Table of the Binomial Distribution*, Washington (Army Security Agency) 1946, pp. iv & 306. Mimeographed.

[33] U. S. Army Service Forces, Ordnance Department, Industrial Service, *Ordnance Inspection Handbook, Standard Inspection Procedures, Quality Control*, Washington (Army Service Forces, Ord-M608-8, Issue 2) 1944, pp. 22 & vi. Lithoprinted.

[34] U. S. Army Service Forces, Quartermaster Corps, Inspection Service, *QMC Inspection Handbook on Sampling for Quality Control*, New York (Army Service Forces, QMC-M608-15) 1944, pp. 81. Mimeographed.

[35] U. S. Navy, *General Specifications for Inspection of Material, Appendix X, Standard Sampling Inspection Tables for Inspection by Attributes*, Washington (Government Printing Office) 1946, pp. 176.

[36] U. S. Navy, Material Inspection Service, *Administration Manual. Standard Sampling Inspection Procedures*, Washington (Navy Department, NAVEXOS P-120-D4) 1945, pp. 244.

[37] U. S. Navy, Office of the Assistant Secretary of the Navy, Material Division, Field Operations Branch, *Instruction Guide for Inspectors of Items. Standard Sampling Inspection Procedures (SSIP) by Attributes*, Washington (Navy Department, NAVEXOS P-471) 1947, pp. iv & 41.

[38] Wald, A., "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, 16 (1945), 117-36. (Reprints available from Institute of Mathematical Statistics, 116 Rackham Hall, Ann Arbor, Mich.)

[39] Wald, A., and Wolfowitz, J., "Sampling Inspection Plans for Continuous Production which Include a Prescribed Limit on the Outgoing Quality," *Annals of Mathematical Statistics*, 16 (1945), 30-49.

[40] Welch, B. L., "Specification of Rules for Rejecting Too Variable a Product, with Particular Reference to an Electric Lamp," *Supplement to the Journal of the Royal Statistical Society*, 3 (1936), 29-48.

[41] Working, Holbrook, *A Guide to the Utilization of the Binomial and Poisson Distributions in Industrial Quality Control*, Stanford, California (Stanford University Press) 1943, pp. 15. Lithoprinted.

Résumé

L'inspection par échantillonnage est la méthode d'accepter ou de rejeter des lots de matériel suivant la qualité d'échantillons. Un plan d'inspection par échantillonnage consiste à spécifier des règles pour: I) former un lot d'inspection, II) sélectionner un ou plusieurs échantillons du lot, III) déterminer la qualité de l'échantillon ou des échantillons, et IV) accepter ou rejeter le lot dans son ensemble suivant la qualité de l'échantillon (des échantillons). Une procédure standard d'inspection par échantillonnage est un ensemble de directives pour la sélection et l'opération de plans d'inspection par échantillonnage.

Dans la plupart des plans d'inspection par échantillonnage, la qualité du lot est mesurée par la proportion de pièces ne répondant pas aux caractéristiques imposées et la qualité de l'échantillon est déterminée en classifiant les pièces de l'échantillon comme acceptables ou défectueuses, sans mesurer le degré d'acceptabilité ou de défectuosité. Cependant, parfois des mesures plus compliquées de la qualité des lots ou des échantillons sont utilisées.

L'échantillonnange simple consiste en l'examen d'un nombre fixe de pièces pour chaque lot; la Méthode de score ("sequential sampling") consiste en l'examen d'un nombre de pièces dépendant des résultats successifs obtenus en cours de travail. La méthode de score comprend l'échantillonnage simple réduit ("curtailed simple sampling"), l'échantillonnage double, l'échantillonnage multiple, et l'échantillonnage pièce par pièce proprement dit. L'échantillonnage simple minimise le nombre maximum de pièces examinées d'un lot quelconque, le nombre d'échantillons individuels d'un lot quelconque, et la varabilité parmi les lots

du nombre de pièces examinées; la méthode de score minimise le nombre moyen de pièces examinées par lot.

Les conséquences statistiques les plus importantes d'un plan d'échantillonnage sont exprimées par sa courbe caractéristique (ou courbe OC). Elle montre, en fonction de la qualité du lot, la probabilité que l'échantillonnage au hasard résultera en un échantillon (ou série d'échantillons) répondant au critère d'acceptation. La courbe OC mesure le degré de protection du destinataire, grâce au plan d'échantillonnage, contre l'acceptation de matérial défectueux qui lui a été offert. Elle montre au producteur les exigences imposées à la qualité et, pour cette raison, influence la qualité produite. De cette manière, l'inspection par échantillonnage a une importante fonction administrative, en coordonnant les rôles des différents intéressés; c'est principalement cette fonction administrative qui fait distinguer sous des rapports fondamentaux, l'inspection par échantillonnage d'autres tests de signification.

Une procédure standard pour la sélection et l'opération de plans d'échantillonnage décide, sur la base de l'expérience générale, beaucoup de questions qui pourraient être décidées par un expert seulement après examen soigneux de la situation particulière. Pour d'autres questions, l'intervalle des choix possibles est réduit à un nombre relativement restreint d'alternatives, avec des instructions détaillées pour faire le choix parmi elles. Et certaines questions sont divisées en parties composantes où, pour chaque composante, la décision peut être prise par un expert en la matière. Bien que la standardisation de méthodes d'inspection par échantillonnage peut avoir pour résultat l'emploi de plans qui ne sont pas aussi bons que ceux qui pourraient être établis par des experts après examen approfondi de situations particulières, la standardisation rend possible d'introduire des plans utiles d'inspection par échantillonnage d'une manière rapide et économique.

La base d'une procédure standard d'inspection par échantillonnage est une collection systématique de plans d'échantillonnage, en même temps que des renseignements relatifs à chaque plan. Une procédure standard énumère les questions à répondre pour faire de la collection la sélection du plan le plus adéquat au problème envisagé; et elle assigne les responsabilités et sert de guide pour la réponse des questions. Les questions doivent être telles que l'effet d'une certaine réponse est le même, quelles que soient les réponses données aux autres questions; ceci est nécessaire pour une division effective de la responsabilité administrative.

THE APPLICATION OF STATISTICAL THEORY TO THE PREPARATION OF INDUSTRIAL SPECIFICATION AND ACCEPTANCE PROCEDURES

by John H. Curtiss
Chief, Applied Mathematics Laboratories
National Bureau of Standards (United States)

In this paper, the word *Specification* denotes a document containing (1) a more or less exact description of those characteristics which render a given manufactured item or a given basic material acceptable, and (2), instructions as to the methods to be used to obtain evidence that the item or material really has the characteristics of acceptability. The specification is made available to the producer or supplier of the item or material when he is invited to bid on a contract or order, and if his bid is accepted, the specification is made a part of the contract.

Such specifications are now used extensively in large-scale procurement in the United States, and particularly in Government procurement. Three especially noteworthy series of specifications are the basic materal specifications promulgated by the American Society for Testing Materials, the miscellaneous purchase specifications issued by the Federal Specification Board (known as Federal Specifications and catalogued in the Federal Standard Stock Catalogue), and the somewhat more specialized and exacting series of purchase specifications of the Navy Department.

The fundamental distinction between direct statements as to the required properties of the item or material, and statements concerning the amount and nature of evidence deemed necessary to establish that required properties are present, is now widely recognized by writers of specifications, and generally the two types of statement are carefully separated in the editorial format. We shall call statements of the first type, *design requirements*, and statements of the second type, *acceptance requirements*. Acceptance requirements fall into two further classes:

(a) *Sampling requirements* (including as a special case the requirement of 100 percent inspection). These comprise the administrative procedure whereby the material is accepted or rejected.

(b) *Methods of test.* These consist of the detailed procedures for obtaining the data on which the administrative procedure is based.

In most series of specifications, it is customary to include design requirements and acceptance requirements in the same leaflet. It should be mentioned in passing, however, that there is a school of thought in the United States which holds that the two types of requirement should always be kept separate, to the end that the acceptance requirements corresponding to a given set of design requirements can readily be tailored to fit local circumstances and the changing economic picture.

In establishing design requirements for specifications, the central statistical technique is that of estimation in the Fisherian sense. Sampling requirements rest most frequently on tests of parametric statistical hypotheses, although considerable use has been made of the theory of a certain stochastic process identified in quality control work with the concept of the average outgoing quality. Various modern statistical techniques play significant roles in the development of methods of test; perhaps the most pressing practical problem here is the prediction of service performance from laboratory tests, so regression analysis with the associated predictive and inferential techniques is of central importance. The statistical aspects of the development of methods of test will not be discussed further in this paper. The establishment of design requirements by statistical methods and the setting up of statistically based sampling plans in specifications will now be discussed in some detail.

In establishing quantitative design requirements, the main engineering problem is to set up realistic tolerance limits. This leads to a statistical problem which, in its simplest formulation, consists of that of estimating one or two of the extreme percentiles of a frequency distribution from sample data representing the distribution. An alternative formulation, which may or may not lead to the same estimators, depending upon the mathematical interpretation given to the verb "to estimate," consists in stating that the problem is, *e.g.*, to estimate the central 99% range of the parent frequency distribution. This type of estimation problem has come to be known in the mathematical literature as the problem of tolerance limits. Various solutions have been developed; a typical approach consists in setting up an interval from the sample data with the property that the mean value in repeated sampling of the portion of the parent frequency distribution lying over the interval is, *e.g.*, 99% and the variance of this portion of the parent frequency distribution is less than, say, 0.001.

However, the solution to the statistical problem of tolerance limits is not the whole story in establishing realistic and workable specification tolerances, and there are usually a number of additional considerations of non-statistical nature which must be taken into account. For one thing, practical considerations of procurement render it frequently advisable to set up the specification limits so that they will include the ranges of variability of many manufacturers. Also, a specification must naturally depend heavily on the intended use of the material or item. Thus engineering and economic considerations must be the final determining factor in the preparation of a specification.

It has been found that the following set of rules provides a useful guide to the solution of the problem of setting up specification tolerance limits for a characteristic such as a dimension, or a physical property, or an element of a chemical composition. (No pretense to completeness is made here, and the circumstances of different cases necessitate many deviations from the formula.)

(a) Obtain test data for material from as many of the prospective sources of supply as possible.

(b) Perform tests of statistical hypotheses to determine whether or not the frequency distributions of the different sources of supply are significantly different. (If for example, the dispersions of all the sources of supply are not significantly different, then all the data for all sources can usually be pooled together for a more accurate estimate of a hypothesized common dispersion.)

(c) Set up a solution of the statistical problem of tolerance limits for each statistically distinct source of supply.

(d) By synthesizing the information obtained in (c) with knowledge as to the engineering requirements of the material and the procurement situation, establish allowable extreme values for the statistical tolerance limits of any possible source of supply.

In performing the steps outlined above, it is possible in certain cases to use laboratory data from research projects in which the primary object was originally to compare one material or treatment with another. However, a satisfactory test of the type of statistical hypotheses involved in research projects often requires a smaller amount of data than the amount needed to estimate population parameters with the accuracy necessary for specification purposes. Moreover, when the results of research are to be used as a practical yardstick which will be used day in and day out to measure the quality of purchased material, various manufacturing testing variables are sure to enter into the picture which were not entirely foreseen in the research stage. It is well to emphasize in this connection that when a specification limit is being set up for some physical or chemical characteristic of a material, the population which must be studied is the population of apparent values of the characteristic as revealed by test data. Thus the dispersion of the population must include the component due to the error (or lack of precision) of the test, which in routine inspection is likely to be greater than in a laboratory experiment. Therefore in many cases the only practicable way of setting up tolerance limits in accordance with the rules given above consists in taking sample units from the current production of several reliable manufacturers and putting these units through the proposed tests in a manner which seems to be representative of the way in which routine inspection tests will be conducted.

Statistically based tolerance limits derived from routine inspection test data were written into a large number of Navy Department purchase specifications during the war. Some of the materials covered by these specifications were: heat treated glass, fiber glass, resistance wire, electric cable, electrical insulation of various types, chemical compounds, refractory materials, rubber products, meters and measuring devices, insecticide dispensers, wood products, rope and yarn, various plastics, life saving devices, electronic equipment, and many different metallurgical materials and products.

The problem of setting up sampling plans for specifications will now be discussed. The theory and practice of statistically based industrial sampling plans has been treated extensively in the literature and only the aspects of industrial sampling most closely related to the writing of specifications will be touched upon here.

When sampling requirements are written into a specification which is supposed to have some measure of permanence, the chief problem of course is to prepare instructions which can be followed under a wide variety of procurement situations and inspection arrangements, and which will at the same time be sufficiently succinct and simple so that field inspectors and contractors will not be confused and antagonized. The older method of dealing with the problem consisted in omitting entirely any instructions as regards sampling for visual inspection (that is, surface inspection of individual units for obvious flaws, gauging of dimensions of units, weighing of units), and setting up very weak sampling plans for laboratory tests. As examples of such plans the following are typical : (a) Test one unit in each sub-lot of 250; if this, unit fails, reject only that particular sub-lot; (b) Subject two sample boiler tubes to a certain physical test ; if any fail, subject two more to the test, and accept the lot only if the latter two pass the test. It should be added that the definition of lot was usually very vague in the older specifications, and direct instructions as to when to reject were often absent.

On the other hand, in defense of these weak plans, three arguments can be brought forward: (1) Material presented for inspection under a specification containing a stable and well-established set of design requirements is typically either almost 100% in conformance with these requirements, or there is a heavy percentage of non-conforming material (that is, the *a priori* distribution of the true percent defective tends to be U-shaped); (2) Purposive rather than random selection of samples was often intended and sometimes was actually practicable (the older editions of the Navy's general instructions to inspectors directed that samples should always be chosen so as to represent the poorest part of the material offered for delivery); (3) Lack of precision in methods of test tends to increase the producer's risk at the high quality levels, so that the contractor should be allowed the benefit of doubt as much as possible.

Since laboratory tests are frequently expensive and destructive, the above arguments suggest that it is appropriate in general to set up for such tests sampling plans which employ low sample sizes and (if the sampling is by attributes) zero acceptance numbers. Plans which are as weak as the examples cited above are not to be recommended, however, except in the rare case in which the authorized representatives of vendors and purchasers, in full awareness of the operating characteristic of such a plan, have come to a rationally based agreement that the plan is appropriate. In the case of single sampling by attributes, sample sizes of from two to fifteen, with acceptance number of zero or one, will often be found to be satisfactory. It is of course assumed that each plan is rationally selected by combining knowledge of its operating characteristic or average outgoing quality limit with knowledge as to the nature of the tests and the quality history of various key suppliers. In the case of a non-zero acceptance number, the equivalent double or sequential plan can conveniently be written out in the specification instead of the single sampling plan implied by the foregoing discussion, but this should be done only with first-hand knowledge of probable inspection arrangements. The entire inspection procedure should be carefully supplied with the sampling plan: the definition of a sampling lot for purposes of the test should clearly be given, the method of sampling (whether purposive or random) should be stated, the order of performing tests on the sample units should be specified, and rejection provisions should be explicitly written out.

The situation is rather different with visual or surface inspection, which, as was previously stated, has been generally neglected in writing sampling requirements for specifications. Here intermediate percents defective are often encountered, so that *a priori* distributions may not be U-shaped; and indeed, for some types of surface defects an intermediate percent defective may be considered entirely tolerable. Also, purposive sampling can seldom be used. These considerations militate against the use of simple, easily written out sampling plans with small sample sizes and low acceptance numbers. Yet experience proves that if inspectors and contractors are given no instructions at all, haphazard and ineffective inspection will surely be the rule rather than the exception.

A satisfactory solution seems to be as follows: Assume sampling by attributes. Prepare as part of the specification, or as a supplement thereto, an exhaustive catalogue of visual defects, classified as major or minor. Publish in the specification for each classification either an "acceptable quality level" (*AQL*) in terms of true long-term percent defective (that is, a percent defective which should entail a sampling rejection only a nominal proportion of times, say one time in twenty); or, if 100% inspection of rejected lots is feasible, publish an average outgoing quality limit (*AOQL*) for each class of defects. Prepare and place in the hands of the inspectors a compendium of attributes sampling plans, arranged primarily according to *AQL* and *AOQL*, and secondarily according to lot size, severity of inspection, and other factors. In a series of specifications, this compendium should be given the status of a standard appendix to each individual specification. The understanding then is that the inspector will select from the compendium what he considers to be an appropriate plan with the specified *AQL* or *AOQL*. The plan will involve heavy or light sampling in accordance with the quality history of the individual supplier; the compendium should contain objective criteria to guide the inspector in his choice of plans within *AQL* or *AOQL* groups.

With the publication of Appendix X of the Navy Department's General Specifications for the Inspection of Naval Material, a system of incorporating visual inspection provisions into Navy Department specifications similar to the one outlined above was inaugurated. Appendix X appeared after the termination of World War II, and sinceac tivity in the preparation and revision of Navy Department Specifications has not been carried on at a high level in the postwar period, there has been no extensive experience with the system. During the war, explicit rationally based sampling plans for both laboratory tests and for surface inspection were incorporated into over 1000 Navy Department Specifications; the large majority of the plans were statistically based.

In conclusion, it should be stated that the operation of acceptance sampling plans is always made more successful by the exhibition and study of the inspection data on Shewhart control charts. The applicability of this statement to sampling requirements in specifications is borne out by extensive wartime experience with the use of such charts in connection with specifications for insecticide dispensers, pneumatic life preservers, mess gear, and many other items of equipment.

Résumé

L'auteur distingue deux types fondamentaux d'énoncés rencontrés dans les spécifications pour produits manufacturés et matières premières, à savoir: les *conditions de plan* (design requirements) et les

conditions d'acceptation (acceptance requirements). Ces dernières se subdivisent en *conditions d'échantillonnage* (sampling requirements) et *méthodes d'inspection* (methods of test).

Les techniques statistiques principales pour l'établissement des conditions de plan et des conditions d'échantillonnage (mais non pour l'élaboration des méthodes d'inspection) sont ensuite discutées.

Le procédé statistique principal pour l'établissement des conditions de plan est l'estimation dans le sens de Fisher, particulièrement des percentiles et des intervalles centraux d'une loi de distribution (le problème des limites de tolérance dans la littérature statistique). Cependant, la solution théorique de ces problèmes d'estimation n'est pas suffisante en pratique et doit être modifiée par l'introduction de considérations technologiques et de faits résultant des possibilités d'achat.

Dans l'élaboration des conditions d'échantillonnage, deux problèmes se présentent: celui des plans d'échantillonnage (sampling plans) pour l'inspection destructive ou coûteuse, et celui des plans d'échantillonnage pour l'inspection visuelle non-destructive. Dans le premier cas, il est avantageux d'employer de petits échantillons et—pour des échantillonnages par attributs"—de très petits nombres d'acceptations (acceptance numbers). Dans le second cas, d'autre part, il est préférable de préparer et mettre à la disposition d'inspecteurs une collection de plans d'échantillonnage auxquels les spécifications se réfèrent en donnant des *niveaux de qualité acceptable* (acceptable quality levels) ou des *limites pour la qualité moyenne de l'article* (average outgoing quality limits), pour les divers types de défectuosités. Dans l'un et dans l'autre cas, des instructions précises pour la méthode d'inspection et l'action qui doit en résulter, doivent être formulées à l'usage des inspecteurs.

Discussion

Mr. Delaporte noted that the usual literature on quality control assumes incorrectly that quality is always normally distributed.

Mr. W. R. Thompson indicated that he has investigated statistical methods not involving assumption of the normal distribution, and has developed the general expressions for evaluating any percentile and interval on which one may express a confidence (a definite probability) that the percentile lies in that interval.

THE DEPENDENCE OF EARNINGS ON CHOICE OF SPECIFICATIONS

by Merrill M. Flood
Assistant Deputy Director of Research and Development
General Staff, United States Army

INTRODUCTION

Earnings of a company depend upon many factors of management and chance. The choice of products, and their detailed specifications, is an important management problem. This problem is outlined in the present paper from the point of view of the statistical engineer. The discussion is referred to an example taken from the glass container industry. The remarks are easily translated to apply to other commodities or services. The object of this brief note is to indicate the principal points at which modern methods of analysis are applicable in the solution of a broad problem of industrial management.

The estimation of the likely effect, on company earnings, of a proposed minor change in specifications, for one of its standard products, is the principal problem discussed. Techniques that are suitable for this estimation problem can usually be applied to the more difficult estimation problems associated with the effects of major changes in specifications, or with the substitution or addition of new products in the line of commodities placed on the market by the company.

Three broadly different methods of approach to the problem are available:

a) *A priori.* Theoretical estimation of demand for the product and of costs for engineering, production, distribution, financing, and management.

b) *Experiential.* Statistical analysis of past earnings of various companies with closely related specifications.

c) *Experimental.* Limited trial of the change and observation of the effect on earnings.

These same three methods, and combinations of them, can be considered for use in nearly any situation where the likely effect of any action is to be estimated. They are discussed in turn, and are made concrete by an example taken from the glass container industry.

EXAMPLE

The example concerns a manufacturer of glass tumblers who serves the hotel and restaurant trade, and operates under procedures long established for the production and distribution of a tumbler that is the company's only important product. It is assumed that the company has under consideration a proposal to alter a physical dimension of the tumbler, and desires an estimate of the likely effect on its earnings of the proposed change in specifications. The only characteristic of the tumbler that is of real importance to the trade, other than the selling price, is durability as indicated by the number of times the tumbler can normally be used before it breaks and must be replaced.

A PRIORI METHOD

The *a priori* method starts with the preparation of detailed plans for the changes necessary in procedures for production, distribution, financing, and management. Cost and time estimates are then made for the engineering work necessary to complete the development and testing of the new product, and for all other important procedural changes. Finally, demand is estimated, market price is established on the basi of cost estimates and planned production schedules, and earnings are estimated.

Statistical methods may well be applied during the planning and estimating stage in connection with the following:

a) Design of laboratory experiments anticipated for the development stage.

b) Design of laboratory and field tests that will be necessary to measure the likely performance of the new product.

c) Design of plans for process control, sampling inspection, and lot acceptance based on the pertinent altered properties of the product that will result from a successful development.

d) Estimation of likely output characteristics of the altered production process, such as: percentage rework and rejects, and average outgoing quality level.

e) Design of advertising tests, and market investigations, that are to be conducted prior to the actual sales campaign for the new product, if development is successful.

f) Design and interpretation of tests to measure the probable demand for the new product.

g) Design of packaging tests, both for sales and shipping purposes, including tests to measure likely damage during production, distribution, and storage.

h) Analyses to determine optimum lot size, capital and facility requirements, and optimum values of other management factors that depend upon performance characteristics of the product or that vary with other changes in the production or distribution processes.

This list indicates the complexity of the problem if it is approached from first principles. In practice, the *a priori* method is seldom used, even in part. It is discussed in this paper in order to provide a partial enumeration of the factors underlying the main problem.

EXPERIENTIAL METHOD

The experiential method starts with the collection of data on earnings, specifications, and pertinent management factors for several manufacturers, or for one manufacturer during several different periods, who have produced and marketed items with specifications closely related to those in question. The successful application of this method requires that a sufficiently accurate relationship be established between earnings and a battery of related factors, including the specification variables under consideration. A set of appropriate values representing each product is then substituted for the dependent variables in the relationship and earnings are thus estimated for each alternative.

The method is seldom useful when applied to data for several companies, may be subject to serious error arising from general business fluctuations when a single company is analyzed over different time periods. It is most useful for a single company that has itself in the past offered a choice of similar products manufactured to slightly different specifications.

An important example of the latter type is provided by an article that varies, in normal production, with respect to the specification under consideration. This case is often met, as for the tumbler example, with particular reference to the problem of determining the relationship between product durability and alternate choices of specifications. Use may be made of the principle that the stable renewal rate is inversely proportional to the durability of the product, thus making it possible easily to estimate relative durabilities as between products manufactured to different specifications. This is accomplished by comparing the frequency with which each type is supplied with the frequency with which it is found in use.

EXPERIMENTAL METHOD

The experimental method starts with the preparation of a plan for a test designed to provide direct evidence of the effect on earnings of a particular change in specifications. The initial planning work is then done, as for the *a priori* method, but estimates of costs, demand, and earnings are eventually derived from data gathered during the test. The test data are interpreted in the same manner as in the experiential method.

The experimental method has the advantage that important variables can be controlled, and the collection of data can be arranged so as to provide observations in numbers adequate to yield estimates as precise as may be desired. The method is apt to be quite costly and very slow.

The importance of careful and efficient test design can hardly be over-emphasized. Recent advances in the art and science of test planning provide important new tools for the analyst. For example, the cost and time for a test can sometimes be reduced substantially by a sequential variation of the arrangement as the experiment progresses.

CONCLUSIONS

It is concluded that careful analysis is usually required in order to estimate the magnitude and direction of effect on earnings of a particular change in specifications. As a working procedure for such an analysis it is usually desirable to start with a rough *a priori* study, next to check and strengthen this study with the help of data readily available from past experience, and finally to conduct an experiment designed to provide new information of the type found to be most needed as a result of the preliminary analyses.

Résumé

Le montant des bénéfices tirés d'une entreprise dépend de nombre de facteurs aléatoires et de facteurs propres à l'exploitation. Le choix de produits à manufacturer et la spécification de leurs caractéristiques présentent des facteurs importants. L'auteur esquisse les problèmes y relatifs du point de vue statistique. Il se base sur un exemple pris dans l'industrie des récipients en verre, mais les observations peuvent aisément s'appliquer à d'autres articles ou services. Ce bref mémoire a simplement pour objet de signaler les points où les méthodes modernes d'analyse peuvent être appliquée à un problème très vaste de l'exploitation d'entreprises industrielles.

Le point qui fait l'objet principal de la discussion est l'évaluation des effets probables, sur le montant des bénéfices, en cas d'un léger changement dans les spécifications d'un des produits-types. La technique applicable dans ce cas pourra généralement être utilisée aussi

dans des cas plus difficiles où il est question de modifications plus importantes, ou de la substitution d'un article ou de l'addition d'articles nouveaux à ceux mis sur le marché par l'entreprise.

Trois méthodes différentes peuvent être utilisées :

a) *a priori*: estimation théorique de la demande du produit en question et des frais de production, de distribution etc.

b) *d'experiénce*: Analyse statistique des bénéfices réalisés antérieurement par différentes entreprises manufacturant des produits semblables.

c) *expérimentale*: Essai du changement sur une échelle limitée, et analyse de son influence sur les bénéfices.

L'application de ces trois méthodes, séparément ou en combinaison, peut être envisagée dans presque toutes les circonstances où il s'agit d'estimer les conséquences probables d'une action quelconque. Dans ce mémoire, chacune des trois méthodes est discutée et illustrée d'un exemple pris dans l'industrie des récipients en verre.

L'auteur arrive à la conclusion qu'en général une analyse approfondie est nécessaire pour l'estimation qualitative et quantitative des effets d'un certain changement sur les bénéfices. Pour une telle anaa e,iillys sernd iqué ordinairement de commencer par une étude *a priori* approximative, dont il faudra vérifier et confirmer les résultats à l'aide des chiffres qu'a fourni l'expérience antérieure; enfin, on aura recours à une expérimentation en vue d'obtenir les renseignements qui, suivant le résultat des analyses préliminaires, ont prouvé être nécessaires.

STATISTICAL METHODOLOGY SECTION

(WITH THE ECONOMETRIC SOCIETY)

Thursday, September 11, at 9:30 a.m.

Statistical Inference

CHAIRMAN :

A. Wald

RAPPORT SUR UNE ENQUETE INTERNATIONALE RELATIVE A L'ESTIMATION STATISTIQUE DES PARAMETRES

Par Maurice Frechet

Faculté des Sciences de Paris (France)

TEXTE DE LA CIRCULAIRE ENVOYÉE

De nombreuses controverses ont eu lieu récemment autour des méthodes d'estimation des paramètres d'une loi de probabilité.

Ces méthodes concernent le problèmo fondamental de la statistique, celui des sondages en vue des prévisions, celui consistant à conclure de la partie au tout.

Il paraît donc opportun et même urgent de confronter les diverses opinions et de faciliter leur comparaison en les concentrant sur un seul exemple.

Le Bureau de l'Institut International de Statistique a décidé de provoquer une enquête, ouverte à tous, sur ce sujet.

Choix du problème.—Dans l'exposé des méthodes d'estimation, ou dans leurs applications à un problème déterminé, les raisonnements présentés sont souvent obscurcis par l'intervention de calculs mathématiques complexes nécessités par les examples traités, calculs qui pourtant ne touchent pas au fond des questions en litige, lesquelles ne relèvent que des principes du Calcul des Probabilités et du sens commun. On éliminera donc celles des considérations mathématiques qui nesont qu'accessoires en se limitant à un problème particulièrement simple.

Question posée.—Un événement E a été observé r fois dans n épreuves. Sachant (ou admettant) que E y avait une probabilité constante (mais inconnue) p, que peut-on dire sur la valeur inconnue de p, connaissant seulement n et le nombre r de répétitions de E (ou sa fréquence $f = r/n$)?

A-t-on des objections (et lesquelles?) à faire contre l'une ou l'autre des méthodes connues qui donnent une réponse à cette question?

Pour simplifier à la fois la réponse et la comparaison des réponses, on ne demande d'examiner que le cas où $n = 1$ et $r = 1$.

[Il importe peu que; vu le petit nombre d'épreuves envisagé (une seul!), la réponse ne puisse restreindre que très peu (et même d'une façon inefficace en pratique) l'incertitude sur la valeur de p. Seules la possibilité et la justification de cette restriction sont en cause ici.]

INTRODUCTION

Il existe deux catégories de personnes s'occupant de calcul des probabilités ou de statistique mathématique. Dans la première nous plaçons ceux qui ont fait de ces sujets leur occupation principale ou l'une de leurs occupations principales. Dans la seconde, nous rangeons ceux pour qui, très légitimement, ces deux sciences ne sont que des moyens en vue de leur utilisation pour une technique ou une science tout autre. Ces derniers ont besoin de recettes toutes faites, ils doivent pouvoir accepter celles-ci comme on accepte l'ordonnance d'un médecin. Il importe alors que les membres de la première catégorie ne donnent pas en toute occasion, à ceux de la seconde, l'impression de sécurité à laquelle conduit souvent l'emploi des mathématiques, sécurité qui existe, en effet, mais seulement du passage des hypothèses aux conclusions, mais qui cesse aux deux extrémités. Or, très souvent, cette réserve n'est pas mise suffisamment en lumière.

En particulier, les biologistes, les psychologues, les ingénieurs, etc., qui usent des méthodes d'éstimation statistique des paramètres ne sont pas toujours très bien informés par les manuels de statistique mathématique de la signification exacte des tests de signification, des intervalles de confiance et surtout des coefficients de confiance et des probabilités fiduciaires. En outre, ils ne savent pas qu'il y a entre les mathématiciens de graves divergences sur ces sujets, qu'il n'y a même pas deux écoles tranchées, mais une variété infinie d'opinions.

Il nous a donc semblé utile de porter ces divergences à la connaissance directe de tous les membres de l'Institut International de Statistique, par le moyen d'une enquête.

"Le problème que vous posez,—nous écrit M. Ville—bien qu'ayant une allure paradoxale,[1] est en effet le problème central du calcul des probabilités. Si on ne peut résoudre ce problème, on ne peut en résoudre aucun... Je dois vous avouer ma crainte que l'on ne puisse obtenir une réponse définitive; mais si vaus parvenez à faire admettre que le problème tel qu'il est posé, et pas un autre, est bien le problème crucial, le résultat sera loin d'être négligeable."

---

[1] Il s'agit évidemment de la réduction à l'unité du nombre des épreuves considérés dans l'enquête.

Nous partageons la même crainte que M. Ville, mais nous avons l'espoir que cette enquête contribuera tout au moins à rendre plus claires et plus précises les objections aux différentes méthodes et surtout à leurs interprétations, et, ce faisant, à éliminer quelques unes de ces objections et à rapprocher les points de vue. D'autant que ces résultats, qui seront, selon nous, atteints dans ce rapport par la simple confrontation des doctrines, pourront être poussés plus loin grâce à la discussion à laquelle il donnera lieu à la session de Washington.

Nous sommes heureux d'avoir vu le Bureau de notre Institut accepter ce point de vue et nous le remercions de nous avoir fait l'honneur de nous charger du rapport sur cette enquête.

Nous ne pouvions naturellement citer dans ce rapport que des extraits des réponses qui nous sont parvenues. Aussi, recommandons-nous au lecteur de prendre connaissance des textes *in extenso* des réponses qui suivent ce rapport et dont celui-ci n'est qu'une préface.

Signalons, en outre, que des questions analogues ont été traitées à un colloque organisé à Atlantic City (U.S.A.), en Janvier 1947, par la branche américaine de la Société économétrique, par l'Institut (américain) de Statistique mathématique et par l'Association américaine de Statistique.

D'autre part, rappelons que M. Gini a fait, sur la question de l'estimation des paramètres, ainsi que sur d'autres questions fondamentales de la statistique un certain nombre de conférences publiées par la Société Statistique italienne, conférences pleines d'exemples suggestifs et de réflexions utiles à méditer sinon à adopter entièrement (voir plus loin), mais que nous ne pourrions résumer ici. Elles ont été fondues par l'auteur en un exposé systématique sous le titre: "*Intorno alle basi logiche e alla portata gnoseologica del metodo statistico*" (*Statistica*, 1945/46).

Enfin, M. Kolmogoroff[2] a publié pendant la guerre un très important mémoire sur l'estimation des paramètres, mémoire que nous n'avons pas eu en mains au moment de la rédaction de ce rapport mais qui semble présenter entre les thèses qui s'affrontent le point de vue intermédiaire auquel il faudra probablement se rallier (voir plus loin).

---

[2] "Sur l'estimation statistique des paramètres de la loi de Gauss," *Bull. Acad, Sciences U.R.S.S.*, série math. 6 (1942), pp. 3–32.

LEX AXIOMES ET LEUR INTERPRÉTATION

Pour traiter clairement le problème de l'estimation, il est essentiel d'établir une distinction entre la théorie axiomatique et mathématique de la probabilité d'une part, son application à la réalité de l'autre.

Tous les mathématiciens s'accordent sur la théorie mathématique de la probabilité. S'ils diffèrent sur la définition de la probabilité, c'est en tant que celle-ci fournit une interprétation applicable au monde sensible. Ceci étant, nous pourrons affirmer, en particulier que la connaissance de la fréquence $f = r/n$ d'un événement E de probabilité p dans n épreuves ne fournit à elle seule, et ne peut fournir, *aucun renseignement* sur la valeur de cette probabilité p (et ceci quelle que soit la valeur considérée, petite ou grande, mais finie, de n) *quand on s'en tient à la théorie axiomatique et mathématique de la probabilité.*

M. Borel a évalué à $1/10^{1000}$ une probabilité négligeable à l'échelle cosmique. Il s'agit ici d'une application à la nature, de la notion mathématique de probabilité. Mais si nous sommes tous d'accord (avec M. Borel) pour ne tenir aucun compte pratique d'une éventualité ayant une telle probabilité, par contre, absolument rien ne nous autorise à en faire autant dans la théorie mathématique.[3] A plus forte raison, en sera-t-il de même pour les fréquences. Qu'un événement de probabilité déterminée se répéte 10^{1000} fois dans 10^{1000} éprevues, rien dans la théorie mathématique n'empèche de supposer qu'en continuant la suit d'épreuves, il cesse de se répéter constamment et que, sa probabilité soit égale, par exemple, à un tiers.

M. von Mises écrit:[4] "la réponse à l'enquête est simple, brève et précise dans le cas où $n = 1$, $r = 1$ ... La réponse est: *Rien.* Aucune inférence ne peut être tirée du résultat d'une épreuve ou, comme on le verra plus loin, d'un petit nombre d'épreuves."

Aussi quand M. Neyman écrit: "L'apparition de E un nombre de fois r au cours de n épreuves indépendantes n'implique rien au sujet de sa probabilité. En fait, cette dernière peut avoir toute valeur de 0 à 1," il précise quil faut ici donner au mot "implique" le sens "peut être démontré mathématiquement."

M. Georges Darmois fait ressortir dans sa réponse que dans le cas $r = 1$, $n = 1$, on peut affirmer que l'événement E est possible.

---

[3] Et sans avoir besoin de l'interroger, nous savons bien que serait d'accord avec nous M. Borel, lui qui a introduit la notion fondamentale de convergence presque certaine (c'est-à-dire ayant lieu, sauf peut-être un cas de probabilité nulle), et a ainsi pu ajouter une précision supplémentaire immense et insoupçonnée au Théorème de Bernoulli.

[4] Nous avons traduit en français les passages que nous désirions citer, des réponses rédigées dans une autre langue.

D'ailleurs nous savons que dans la théorie axiomatique un événement (par exemple le choix d'un point donné sur un segment) peut être possible et de probabilité nulle.

Ainsi, il doit être bien entendu que si un événement E se produit dans une épreuve, la théorie axiomatique ne s'oppose à aucune valeur de la probabilité *a priori* de p, même à la valeur zéro. Comme nous le disions plus haut, il en est exactement de même lorsque l'événement s'est produit 10^{1000} fois sur 10^{1000} épreuves.

Il me parait extrêmement important d'insister sur ce point, bien qu'il soit evident, afin de bien faire ressortir la nécessité de ne pas accepter les méthodes d'estimation comme mathématiquement incontestables.

Pour arriver dans le problème posé à une affirmation mathématiquement démontrable, il faut: soit modifier les hypothèses, soit modifier la question posée, soit faire intervenir, à côté de la théorie axiomatique, une interprétation de la probabilité, interprétation qui peut varier d'un mathématicien à un autre.

Première solution.—La manière la plus simple d'opérer est celle qui est préconisée avec de nombreux auteurs par M. Oskar Anderson: "Quant à loi des grands nombres et à sa conséquence la théorie des intervalles de confiance—, je crois qu'elles ne peuvent être basées que sur le lemme non-mathématique suivant énoncé par Cournot: les événements ayant de petites probabilités se réalisent rarement."

Et M. Féraud: "décidons d'adopter le degré de probabilité $\alpha = 1/10^6$ c'est-à-dire, de considérer que tout ensemble auquel est affecté une probabilité inférieure ou égale $1/10^6$. sera vide de résultat."

C'est la première des formes de raisonnement que nous avons exposées dans notre article de *Portugaliae Mathematica*:[5]

"On peut faire le raisonnement suivant: admettons, comme on le fait souvent, que si p est assez petit—par exemple, inférieur à un millionième—on doit considérer E, avant toute épreuve, comme événement pratiquement impossible. C'est, au fait, la règle que nous appliquons journellement, consciemment ou non, pour ne pas nous soucier de certains dangers dont la probabilité est très petite.

Ceci étant, si la probabilité p de l'événement E était inférieure à un millionième, on jugerait, avant toute épreuve, E comme pratiquement impossible et on agirait en conséquence. Or, après l'épreuve, on a constaté que E s'était produit. Puisqu'on aurait jugé, avant l'épreuve, E comme pratiquement impossible, si p était inférieur à un millionième c'est donc (et c'est ici qu'il convient de se demander si cette conclusion si' impose) qu'on ne peut pratiquement supposer, après l'épreuve,

---

[5] "Fondements des méthodes statistiques d'estimation," *Portugaliae Mathematica*, Vol. 5 (1946), pp. 137-141.

$p < 1/10^6$. Par conséquent, dans ce mode de raisonnement, on doit conclure que p est contenu dans "l'intervalle de confiance" ou "l'intervalle fiduciaire" $\frac{1}{10^6} \leqslant p \leqslant 1$.

D'ailleurs, on se contente souvent de déterminations moins rigoureuses des cas pratiquement impossibles; on considérera souvent ainsi ceux dont la probabilité est inférieure à un millième, un centième. D'une façon générale, si la réalisation d'un événement est considérée comme peu probable quand sa probabilité est inférieure à un certain nombre ω $(0 \leqslant \omega \leqslant 1)$ alors, quand un événement E de probabilité déterminée mais inconnue, p, se réalise dans une épreuve, on considérera comme peu vraisemblable que p ne soit pas dans l'intervalle

$$\omega \leqslant p \leqslant 1 \tag{1}$$

et cet intervalle sera dit relatif au "coéfficient de confiance" $1 - \omega$ ou à la "probabilité fiduciaire" $1 - \omega$.

M. O. Anderson précise: "la limite de confiance diffère nonseulement selon les individus, mais aussi suivant les divers degrés d'importance d'un échec. Par exemple, la probabilité 1/100 d'une pluie imminente est négligée, mail il n'en est nullement de même de celle d'une sérieuse infection typhique ou cholérique."

Nous avions ainsi continué: "Bien entendu, il ne saurait être question de considérer $1 - \omega$, comme la probabilité que p appartienne à cet intervalle, puisque p, par hypothèse, s'il est inconnu, est déterminé. C'est pourquoi, ω étant la probabilité la plus grande qui correspond à ce que nous avons appelé peu "probable," nous nous sommes contentés de dire qu'il est peu "vraisemblable" que p n'appartienne pas à l'intervalle,[6] mais nous n'avons pas dit que cela est peu probable, puisqu'on sous entendrait que cette circonstance a une petite probabilité, alors que, pour le moment, du moins, nous ne saurions pas de quelle probabilité il s'agit."

L'interprétation ci-dessus de l'intervalle de confiance n'est d'ailleurs pas unanimement acceptée.

Tout d'abord, certains statisticiens se refusent à examiner le cas d'une seule épreuve qui, disent-ils, ne peut fournir aucun renseignement utile. Si en effet, comme il est d'usage, ω est un nombre petit, les inégalités[7] qu'on aura adoptées, différeront si peu des inégalités $0 \leqslant p \leqslant 1$ qu'elles n'apporteront aucun renseignement utile. Cela est vrai, mais encore une fois, il s'agit ici surtout d'examiner la validité ou l'interprétation de cette inégalité et non son utilité.

---

[6] *Cf.* "Fondements des méthodes statistiques d'estimation," *op. cit.*

[7] *Cf.* "Fondements des méthodes statistiques d'estimation," *op. cit.*

Or les raisonnements qui peuvent être mis en doute dans le cas $n > 1$ pour conduire à un intervalle de confiance, sont exactement les mêmes que pour $n = 1$ et sont en outre noyés dans des calculs mathématiques qui ne sont pas en discussion ici.

Une autre objection concerne la validité du lemme de Cournot. "La logique me permet pourtant, dit M. Divisia, de penser que, considerer comme pratiquement impossible un coup que les mathématiciens appellent (dans leur terminologie vicieuse) hautement improbable, conduit à de l'absurde. Le problème vrai de la conduite humaine consiste, non pas à nier certaines possibilités mais à mettre en balance les conséquences d'une éventualité quelle, qu'elle soit et l'effort à faire ou l'inconvénient à subir pour l'éviter."

D'autre part, nous avions écrit à ce sujet: "Nous allons donner un exemple des difficultés cachées qu'on rencontre, sans toujours s'en douter, dans ce genre de questions.

Indiquons d'abord un cas sans difficultes cachées: on a une urne contenant 25 boules portant respectivement les 25 lettres de l'alphabet. On en tire sept boules. Placées dans leur ordre de tirage, elles reproduisent le mot MIRACLE. Tout le monde pensera sans même qu'il soit besoin de faire le calcul que le tirage a été truqué: un tel événement était pratiquement impossible. Sa probabilité

$$\frac{1}{25}\,\frac{1}{24}\,\frac{1}{23}\,\frac{1}{22}\,\frac{1}{21}\,\frac{1}{20}\,\frac{1}{19} < \frac{1}{(20)^7} < \frac{1}{10^9}$$

était d'ailleurs, non seulement inférieure à un millionième; mais mille fois plus petite encore.

Mais procédons à un tirage non turqué, on obtient par exemple CDATIPS. Un tel tirage a la même probabilité que celui du mot MIRACLE, elle est $< \frac{1}{10^9}$. Il était donc, lui aussi, pratiquement impossible avant l'épreuve et pourtant il a eu lieu. Voilà le paradoxe.

M. Dumas m'a fait très justement observer que la probablité qu'il faudrait faire intervenir, cest plutôt celle de l'apparition parmi les tirages possibles d'un tirage possédant un caractère spécial: par exemple constituant un mot du dictionnaire comme ici, ou bien présentant les lettres dans un order alphabétique. Cette probabilité est égale à la précédente multipliée par le nombre de tels tirages singuliers. Elle n'est donc pas aussi faible que la précédente.

Mais ce qui provoque l'étonnement c'est que nous nous rendons compte instinctivement, que de tels tirages singuliers sont rares parmi les autres c'est-à-dire que, malgré tout, la probabilité, même ainsi renforcée, reste encore extrêmement faible."

C'étaient là des idées analogues à celles qu'exprime ainsi M. Divisia, "ce qu'il y a d'improbable, ce n'est pas que le mot MIRACLE sorte, c'est que si j'ai annoncé la sortie de ce mot, il sorte précisément. Et il en est de même de CDATIPS ou de tel autre."

D'après M. de Finetti, "il n'y a rien de frappant si un événement très peu probable—(que ce soit par exemple, le mot MIRACLE ou le mot CDATIBS) se réalise et onne peut donc, du fait de sa réalisation conclure l'événement avait une probabilité moins petite, de même il n'y a aucune raison de soupçonner que le tirage ayant donné un résultat quelconque (MIRACLE ou CDATIPS) était truqué, ou en général, influencé par une cause indéterminée."

M. von Mises résoud le paradox de la façon suivante:

"Si le mot MIRACLE a été formé, c'est peut-être que la méthode de tirage a favorisé quelques combinaisons particulières, mais celà peut être aussi qu'aucune irrégularité de ce genre ne s'est présentée, c'est-à-dire qu'a la longue, chaque combinaison de 7 lettres apparaitraît avec la même fréquence $p = 4/10^{10}$. Dans ce dernier cas, la methode de tirage sera dite régulière (*unbiased*). Mais dans ce cas, il est donc tout à fait indifférent que la probabilité soit 1/6 ou $4/10^{10}$ Il n'y a aucun sens à dire qu'un événement de probabilité $1/10^{10}$ est "pratiquement impossible;" si l'on employait cette façon de parler d'une façon régulière on devrait dire qu'aucune combinaison de 7 lettres n'est "pratiquement possible," la signification de $p = 1/10^{10}$ nest rien de plus ni de moins que ceci: dans une suite sufficamment étendu l'événement apparaîtra avec cette fréquence, aucune assertion additionnelle ou modifiant la précédent n'est justifiable."

Si nous comprenons bien MM. de Finetti et von Mises, ils réjettent notre paradoxe en le lemme de Cournot (de même par suite que l'évaluation de M. Borel citée plus haut).

Dans les "Fondements" nous avions proposé des explications de ce paradoxe compatibles avec le lemme de Cournot.

"On peut dire: il est impossible de parler de la probabilité de cet événement, puisque, avant l'épreuve, on n'aurait pas su de quel événement il s'agissait. Le paradoxe provenait de ce qu'on parlait d'un événement non défini d'avance.

"On peut aussi dire, la probabilité de ce tirage, c'est la probabilité d'un événement qui est défini comme étant l'événement qui va se réaliser: cette probabilité est donc égale à l'unité et ne peut être $1/10^9$. Il n'y a plus de paradoxe."

Dans la première forme de raisonnement qui vient d'être proposée, une fois choisie la valeur de ω, comme correspondant à la plus grande valeur de la probabilité d'un événement pratiquement impossible,[8] on affirme quand E s'est réalisé qu'il est pratiquement impossible qu'on n'ait pas $\omega \leqslant p \leqslant 1$. Il n'y a pas de degré dans cette affirmation et elle ne suppose pas la connaissance d'autre renseignement que les valeurs $r = 1$, $n = 1$.

---

[8] On peut tenir compte de l'objection de M. Divisia au lemme de Cournot en disant au lieu "d'événement pratiquement impossible," "événement *traité comme* pratiquement impossible."

La "Surprenance"—Avant de passer au cas où l'on aurait d'autres renseignements, signalons une curieuse notion introduite par M. Charles Jordan dans sa réponse et qui est justifiée par un raisonnement assez analogue au précédent. C'est la notion de "surprenance" (*surprisingness*) du résultat d'une épreuve. Dans n épreuves sont observés successivement r_1 fois l'événement E_1 de probabilité $p_1, \ldots, r_s$ fois l'événement E_s de probabilité p_s, ces événements étant incompatibles et leur système exhaustif ($\Sigma p_k = 1$).

Avant les épreuves, la probabilité qu'on obtienne ce système était

$$\varpi = \frac{n!}{r_1! \ldots r_s!} p_1^{r_1} \ldots p_s^{r_s}$$

soit η, la plus grande des valeurs du second membre quand on remplace les r_k par tous les systèmes possibles de répétitions ($\Sigma r_k = n$). Comme on devait plutôt s'attendre à un système dont la probabilité η est la plus grande et s'attendre d'autant moins au système observé que ϖ est plus petit, on sera surpris si ϖ est petit et d'autant plus surpris que $S = 1$ ϖ/η est grand. C'est cette dernière quantité que M. Jordan appelle la "surprenance" du système des répétitions observées. Dans le cas où $n = 1$ et où l'on appelle E_1 l'événement E, on doit prendre $p_1 = p$ et $\varpi = p$ si E est observé, $\varpi = q$ dans le cas contraire. Donc si $p < q$, $\eta = p$; sinon $\eta = q$. Dès lors quand $p < \frac{1}{2}$, si E a été observé, $S = 1 - p/p = 0$, quand $p < \frac{1}{2}$, $S = 1 - p/q$. Si l'on veut que la réalsation de E apetite surprenance $S < \omega$, il faudra que $p \geqslant \frac{1}{2}$ ou que $1 - p/q < \omega$ où $p > (1-\omega)/2-\omega$. Comme $\frac{1}{2} \geqslant (1-\omega)/(2-\omega)$, il suffit que p soit dans l'intervalle de "non surprise"

$$\frac{1-\omega}{2-\omega} < p \leqslant 1.$$

On voit qu'on arrive à un résultat analogue au précédent, sauf qu'on n'a plus une interprétation aussi simple de la plus petite valeur admise pour p.

NÉCESSITÉ DE RENSEIGNEMENTS A PRIORI

Un grand nombre d'auteurs considèrent pourtant qu'il est impossible de rien affirmer sur p connaissant seulement les valeurs de r det de n. "Pour émettre une conclusion, dit M. von Mises, il est indispensable de se servir d'une information additionelle suffisante, préalable aux épreuves et concernant la structure de l'ensemble d'objets duquel l'objet employé dans les épreuves fut choisi." Et d'après M. Neyman, "quelque conclusion sur la probabilité p serait possible si ... l'on saviat que la probabilité p est elle-même une variable aléatoire. Dans ce cas, la formule de Bayes pourrait être appliquée et le résultat de cette application dépendrait de la connaissance de la fonction de distribution de p." On sait en effet que si $f(p)dp$ est la probabilité élémentaire *a priori* pour que la probabilité aléatoire P soit comprise entre p et $p+dp$, la probabilité $p_1(\alpha)$ pour que $P > \alpha$ quand l'événement E s'est produit, est égale au quotient

$$p_1(\alpha) = \frac{\int_\alpha^1 p\, f(p)dp}{\int_0^1 f(p)dp}.$$

Ainsi, sachant quelle était la loi de probabilité *a priori* de P, c'est-à-dire avant l'épreuve, on sait calculer sa loi de probabilité *a posteriori* quand on sait que E s'est produit.

D'après M de Finetti: "Le problème et les méthodes (d'estimation des paramètres) n'ont de sens qu'en s'appuyant sur le raisonnement de Bayes, même s'il est sous-entendu, déguisé ou refusé.

Mais qu'est-ce que la formule de Bayes?

Ici se place une digression historique.

En employant cette expression dans le texte de l'enquête, nous avons eu le tort de nous conformer à un usage fréquent mais historiquement incorrect. Au sens que nous y attachions, nous aurions dû dire: la formule de Bayes-Laplace. Et c'est bien en ce sens que l'entend M. Neyman puisqu'il exige pour l'appliquer la connaissance de la fonction $f(p)$ de distribution de P. Or comme le rappelle M. von Mises dans sa réponse, Bayes n'a pas inclu dans sa formule "le facteur décisif" $f(p)$. Et M. Molina a écrit[9] que cette fonction "n'apparait ni explicitament ni implicitement dans l'essai de Bayes. Ce fait soulève la question de savoir si Bayes a eu ou non la moindre notion du problème général des causes." Par contre, c'est certainement par inadvertance que M. von Mises attribue à Laplace la même lacune que celle de Bayes. Le cas où toutes les hypothèses (ou causes) n'ont pas la même probabilité est formellement considéré par Laplace, avec la formule générale correspondante, aux pages 198-199 du tome VII de l'édition de 1947 de ses *Oeuvres*. S'il est vrai qu'il a fréquemment employe sa formule dans le cas ou les probabilités *a priori* sont égales, ce n'est nullement parce qu'il considérait cette égalité comme allant de soi, mais uniquement pour simplifier le calcul. C'est ainsi que dans le même volume, appliquant la formule de Bayes-Laplace au calcul de la probabilité pour que la somme des orbites soit comprise entre deux nombres donnés, Laplace a soin de spécifier à deux reprises (pour les planètes, p. 281 et pour les comètes,

---

[9] *Bayes' theorem. An expository presentation.*

p. 335) que son calcul est exact *si* toutes les inclinaisons sont également possibles. Enfin, p. 489, il traite un exemple avec quatre hypothèses inégalement possibles.[10]

Il semble donc que ces passages de Laplace aient échappés à M. R. A. Fisher quand il écrivait " ... car Laplace prend pour allant de soi et sous une forme largement généralisée ce que Bayes avait désiré essayer de postuler dans un cas spécial." [11]

Choix d'une loi a priori

Admettant l'emploi de la formule de Bayes-Laplace, certains proposent (ou essaient) son application avec une loi de probabilité *a priori* bien définie.

Par exemple, M. Eyraud note qu'en proposant que cette loi soit uniforme, c'est-à-dire que $f(p) = 1$, on aurait pour la probabilité élémentaire *a posteriori*, quand E a eu lieu, la valeur $\dfrac{pdp}{\int_0^1 pdp} = 2pdp$. D'où

$p_1(\alpha) = \int_\alpha^1 2pdp = 1 - \alpha^2$. La valeur dominante de la probabilitè alétoire inconnue est alors 1, sa valeur moyenne 2/3, sa valeur équi-probable $\frac{1}{\sqrt{2}}$ (avec $\frac{2}{3} < \frac{1}{\sqrt{2}} < 1$), son écart quadratique moyen $\frac{1}{3\sqrt{2}}$.

Pour que l'erreur commise, en supposant que p est dans un intervalle déterminé, (α, β), ait une probabilité au plus égale à ω, il faut que $\int_\alpha^\beta 2pdp \geqslant 1 - \omega$ où $\beta^2 - \alpha^2 \geqslant 1 - \omega$.

On a intérêt à rendre cette estimation la plus précise possible, c'est-à-dire à réduire la longueur de l'intervalle $\beta - \alpha$. Or, celui-ci doit être $\geqslant \frac{1 - \omega}{\beta + \alpha}$.

Il faut donc rendre maximum $\beta + \alpha \leqslant \beta + \sqrt{\beta^2 - 1 + \omega}$, ce qui a lieu pour $\beta = 1$ et $\alpha - \sqrt{\omega}$. Ainsi l'intervalle de confiance qui était $(\omega, 1)$ dans le premier raisonnement (p. 368), se trouve réduit à l'intervalle plus petit $(\sqrt{\omega}, 1)$. Suivant la remarque générale faite plus loin, on obtient une précision plus grande quand on emploie la formule de Bayes. Mais tout ce raisonnement repose sur l'hypothèse de Bayes à savoir la répartition uniforme des probabilités, hypothèse qui peut être conforme aux faits mais qui généralement ne le sera pas, comme l'a montré M. Gini. Et dans les cas où elle le sera, on n'en aura pas toujours la preuve.

---

[10] S'il résulte de ce qui précède que Laplace (et non Bayes) a explicitement écrit la formule rigoureusement exacte dans tous les cas, la question reste ouverte de savoir si c'est Laplace qui l'a fait le premier.

[11] *Inverse probability* (*Proc. Cambridge Philosophical Soc.*, Vol. XXVI, 193. p. 528) D'ailleurs l'expression probabilité inverse *y* est associée par M. Fisher au système formé par la formule de Bayes-Laplace *et* par le postulat de l'égale probabilité *a priori*, alors que ce postulat est indépendant de l'idée d'inversion.

M. Dumas propose une autre loi *a priori*. Pour diverses raisons qu'il expose dans sa réponse, il choisit une loi "virtuelle" $f(p)$. C'est-à-dire qu'il applique la formule de Bayes-Laplace en prenant la limite du second membre quand on y remplace $f(p)$ par une suite de fonctions $f_n(p)$ qui sont des densités de probabilité, sans qu'il en soit ainsi pour leur limite $f(p)$ qu'il prend égale à $\frac{1}{p(1-p)}$. En prenant $f_n(p) = \frac{A_n}{p(1-p)}$ entre $\frac{1}{n}$ et $1 - \frac{1}{n}$ et 0 en dehors, on trouve pour la densité de probabilité *a posteriori* $\frac{1}{(1-p)L(n-1)}$ et ou la fonction de répartition $\frac{L\frac{1}{1-p}}{L(n-1)}$ $- L\frac{n}{n-1}$. Celle-ci entre $\frac{1}{n}$ et $1 - \frac{1}{n}$, tend vers zéro quand $n \to \infty$ pour $p < 1$ et vers 1 quand $p = 1$. Par conséquent, la probabilité *a posteriori* n'est plus aléatoire (dans le cas examiné où $r = 1$, $n = 1$); elle est égale à 1 (presque certainement). Ses valeurs moyenne, equi-probable et dominante sont égales à 1. Peut-on, comme M. Dumas, en conclure que ce résultat est conforme au bon sens? Autrement dit, si l'on sait seulement qu'une seule épreuve ayant été faite l'événement *E* s'est produit, en conclurait-on comme allant de soi, que cet événement devait certainement se produire?

INCERTITUDE SUR LA LOI A PRIORI

Même les partisans de l'emploi de la formule de Bayes-Laplace doivent reconnaître qu'il est souvent difficile de déterminer exactement les lois de probabilité *a priori*. On peut tourner cette difficulté de l'une des deux manières suivantes.

D'abord en observant que, si l'on ne connait pas exactement la loi de probabilité *a priori*, on a généralement quelques renseignements en dehors des valeurs de *r* et de *n*. D'après M. Georges Darmois, "On sait en général, quelque chose sur l'inconnu que l'on cherche mais ce quelque chose n'est pas aussi précis qu'une loi de probabilité *a priori*. Et sans être très précis, c'est pourtant mieux que rien."

Appuyons par un exemple.

Dans l'application du problème actuel aux essais industriels, on emploie généralement la théorie soit comme si l'on ne savait rien, soit au contraire, comme si l'on connaissait des caractéristiques (moyenne, dispersion, ...) de la distribution des objets pouvant être livrés ou la distribution elle-même. En fait, on ne sait rien de précis, mais on sait qu'on a affaire à des fournisseurs qui ne risquent de livrer pas plus des lampes en carton que des vis en or, et même, qui, non seulement s'écarteront des termes du marché de façon beaucoup moins extravagante, mais, en fait, s'en rapprocheront de très près. Il en résulte que si l'on connaissait la loi de probabilité *a priori* qui correspond à cette situation, on obtiendrait, une fois l'essai effectué avec un résultat favorable, une probabilité *a posteriori* (que la livraison soit bonne) plus grande que la probabilité fiduciaire ou le coefficient de confiance résultant de cet essai, quantités qui sont indépendantes de toute probabilité *a priori*. C'est là une circonstance favorable aux théories indépendantes de la formule de Bayes-Laplace car il vaut mieux rejeter de bons lots comme mauvais (lots que le fournisseur pourra d'ailleurs réexaminer et représenter partiellement) que d'accepter de mauvais lots comme bons.

Nous venons de nous appuyer sur une remarque intuitive. Il est toutefois plus sûr encore de la prouver mathématiquement. C'est ce qu'a fait dans sa réponse, M. Georges Darmois en démontrant que dans le cas ou la probabilité inconnue P de l'événement E a une loi de probabilité *a priori* et où cette loi a une densité de probabilité, la probabilité *a posteriori* $p_1(\alpha)$ que P soit supérieur à un nombre positif α, est—quand dans une épreuve E s'est produit—supérieure à la probabilité *a priori* $p_0(\alpha)$ de l'inégalité $P > \alpha$, l'égalité $p_1(\alpha) = p_0(\alpha)$ ne pouvant avoir lieu que quand ces deux quantités sont nulles.

PROBABILITÉ A PRIORI APPROCHÉE

Revenons à l'incertitude des lois de probabilité *a priori*. Quand nous avons, en dehors des valeurs de r et de n, quelques renseignements sur les épreuves et les événements considérés, il en résultera souvent que ces épreuves et ces événements peuvent être ranges dans une "super-population" où p est une variable aléatoire. Bien que, comme dans l'exemple précédent des essais industriels, nous n'ayions généralement pas les moyens de déduire de nos renseignements sa loi de probabilité *a priori*, celle-ci est déterminée par ces renseignements. Il arrivera alors parfois que ne sachant comment déduire exactement $f(p)$ de ces renseignements, nous puissions enfermer $f(p)$ entre certaines limites $f_1(p), f_2(p)$. C'est le cas (ou plutôt le cas particulier où f_1 et f_2 sont constants) qui est signalé par MM. Ville et Fortet. Appelons C la classe des fonctions comprises entre f_1 et f_2. D'après M. Fortet, on pourra considérer comme vraies pour la loi de probabilité *a posteriori* de P celles des propriétés de cette loi qui sont communes à toutes les fonctions de probabilité *a posteriori* correspondant aux fonctions de probabilité *a priori* comprises entre f_1 et f_2.

Dans le même ordre d'idées, M. von Mises indique un mémoire où il a montré comment, dans beaucoup de cas, quelques information générales sur $f(p)$ permettent d'obtenir des inégalités fournissant des renseignements sur les probabilités *a posteriori.*

Un autre moyen, qui, d'après M. Molina, dans son article cité plus haut, a été déja employé par Laplace, Poincaré, et Edgeworth, est signalé par M. von Mises et M. Fortet. Ce dernier se contente de mentionner que lorsque le nombre d'épreuves est grand, les méthodes d'estimation indépendantes de la formule de Bayes-Laplace donnent asymptotiquement le même résultat que cette formule, sous certaines conditions, d'ailleurs très générales. M. von Mises signale que, pour un grand nombre d'épreuves la probabilité *a posteriori* peut devenir indépendante de la loi de probabilité *a priori* et précise dans sa réponse un des systèmes de conditions suffisantes à cet effet: il considère le cas, vraiment très général, où $f(p)$ est borné dans l'intervalle (0, 1) et est continu et $\neq 0$ pour p égal à un certain nombre α $(0 < \alpha < 1)$. M. von Mises a prouvé que dans ce cas, si maintenant $\frac{r}{n} \to \alpha$ quand $n \to \infty$ et si U_1, U_2, p_1, p_2 sont des nombres tels que

$$\frac{U_1}{p_1 - \alpha} = \frac{U_2}{p_2 - \alpha} = \sqrt{\frac{n}{2\alpha(1-\alpha)}} \tag{2}$$

alors la probabilité *a posteriori* pour que $p_1 < p < p_2$, est asymptotiquement equivalente à l'intégrale de Laplace $\frac{1}{\sqrt{\pi}} \int_{U_1}^{U_2} e^{-u^2} du$ laquelle—comme au reste la condition (2)—est indépendante de la fonction $f(p)$.

Par exemple, ajoute-t-il, si E a eu lieu dans la moitié des épreuves, la probabilité que p tombe entre 0.45 et 0.51 est environ 0.47 quand $n = 1000$ et environ 0.95 dans le cas où $n = 10{,}000$.

Par contre, M. von Mises insiste sur l'impossibilité, dans le cas de petits échantillons, de conclure quoi que ce soit sans connaître la structure de la catégorie d'épreuves où P est choisi. Il indique à titre d'éxemple deux mémoires de 1931 où a été appliquée la théorie des petits échantillons (sans formule de Bayes) et où elle a conduit à des résultats qui sont, selon lui, contraires au bon sens. Nous reviendrons sur ce point plus loin.

OBJECTIONS A L'EMPLOI DE LA FORMULE DE BAYES-LAPLACE

La discussion est ardente entre partisans et adversaires, nous ne dirons pas de la formule de Bayes-Laplace, qui est mathématiquement indiscutable, mais de la possibilité et de l'utilité de son emploi pratique.

Les seconds considèrent que la formule de Bayes-Laplace quoique pouvant servir dans quelques exemples concrets particuliers est généralement inapplicable dans les problèmes statistiques qui se présentent dans la pratique et cela pour plusieurs raisons. Même quand p peut être considéré comme une variable aléatoire, on ne connait pas sa loi de probabilité ou bien p peut être défini dans plusieurs catégories d'épreuves, dans plusieurs populations distinctes, ayant respectivement leurs lois de probabilité, qui peuvent être distinctes et entre lesquelles on ne sait choisir. Plus encore, il arrive même généralement que la probabilité de l'événement considéré, E, a une valeur déterminée dont on ne voit aucune raison de la considérer comme l'une de celles qui se présentent comme la probabilité d'un événement E choisie au hasard dans une famille d'événement ou superpopulation.

Par exemple, Mr. Bartlett "rejette l'usage quantitatif de la formule de Bayes-Laplace dans les problèmes d'estimation parce que les probabilités *a priori* qu'elles contiennent—qu'on les regarde comme subjectives ou objectives—ne sont pas, dans son opinion, des quantités auxquelles des valeurs uniques et connues peuvent être attribuées."

M. Derksen s'associe "aux critiques élevées par M. Fisher ... contre l'usage fait du théorème de Bayes pour la solution du problème de la possibilité dite inverse." Parmi ces critiques figurent celle qui concerne l'égalité des probabilités *a priori*. C'est une hypothèse qui a été en effet, considérée par M. Pearson comme l'hypothèse à adopter quand on ne sait rien sur les probabilités *a priori*. Mais il semble bien que la tendance la plus répandue actuellement parmi les partisans de la formule de Bayes-Laplace consiste à n'avoir recours à cette hypothèse que dans deux cas, d'abord celui où l'on a des raisons de penser que cette hypothèse est conforme aux faits, ensuite le cas où le nombre d'épreuves est grand et ou des théorèmes analogues à celui de M. von Mises rappelé plus haut montrent que cette hypothèse, qui facilite les calculs, ne modifie pas sensiblement les résultats.

OBJECTIONS AUX ESTIMATIONS DIRECTES

D'autre part, si M. von Mises concède que "dans beaucoup de problèmes, il est vrai que p (ou toute quantité déduite de p) ne peut être considérée comme une variable aléatoire," il ajoute: "S'il n'est pas possible d'imaginer que l'objet de l'expérience peut varier de façon à correspondre à des valeurs diverses de p, comment peut-on parler de la chance que p tombe dans quelque intervalle? Et si des assertions sont émises qui contiennent d'autres terms que chance ou probabilité, comme vraisemblance (*likelihood*), probabilité fiduciaire, niveau de confiance ou de signification, la première chose à faire serait de donner clairement des définitions opératoires pour ces termes, ce qui est loin d'avoir encore été fait."

Avec M. von Mises, MM. Dumas et Fortet sont d'avis que ceux qui prétendent ne pas employer les probabilités *a priori* et recommandent des méthodes d'estimation fondées sur la seule connaissance des valeurs de r et de n font en réalité des hypothèses implicites.

Il est seulement regrettable qu'aucun d'eux ne précise les points particuliers de raisonnements où de telles hypothèses cachées se sont glissées.

D'ailleurs nous allons voir comment on peut présenter la théorie des intervalles de confiance *de façon à parer aux objections qu'elle a suscitées*, pourvu qu'en même temps, on se résigne à *abandonner certaines prétentions*.

En effet, ces objections portent d'abord sur le cas où une épreuve étant faite et un résultat déterminé ayant été observé: la réalisation de E, il s'agit d'en conclure quelque chose sur p, sans s'occuper des autres épreuves qu'on pourrait faire, ni de leurs résultats possibles, sans s'occuper de savoir ce que pourrait faire le même observateur une autre fois, (car on doit considérer, par exemple, le cas où il ne pourrait tenter l'épreuve qu'une fois dans sa vie).

Dans ce cas, on ne voit pas du tout comment on pourrait attacher un sens à la probabilité que p, nombre inconnu mais déterminé, soit entre deux nombres donnés. Car comme le dit M. Fisher: "les premiers aûteurs ayant étudié la logique des probabilités ont mis en évidence un point de la plus haute importance, qui est très facilement négligé, à savoir que la probabilité est une qualité non d'un événement particulier en soi, mais d'un événement regardé comme membre typique d'une population bien définie d'événements." (Nous dirions d'ailleurs plutôt "regardé comme pouvant être réalisé ou non dans une population bien définie d'épreuves.") Or dans la première forme du problème posé, l'existence de cette population est exclue.

Il semble bien résulter de ce qui précède que si l'on examine cette forme du problème sans savoir rien d'autre que le nombre d'épreuves et la fréquence de l'événement, alors: ou bien l'on se restreint à la théorie axiomatique et l'on ne peut rigoureusement rien conclure sur p; ou bien malgré les objections qu'il soulève on adopte le lemme de Cournot ou un lemme analogue (comme celui de Charles Jordan concernant la "surprenance") et on trouve que p est dans un intervalle déterminé pratiquement.

On peut donner au problème, une seconde forme: si nous répétons l'épreuve, il pourra arriver que, dans certaines, E ne se produise pas. Le problème résolu par la formule de Bayes-Laplace ne s'occupe pas de ces dernières, il fournit la probabilité que dans la catégorie C d'épreuves où E s'est réalisé, p soit compris entre deux nombres donnés. Et cette catégorie C est extraite de la catégorie d'épreuves où la probabilité *a priori* de p a la valeur qu'on lui a attribuée pour l'utiliser dans la formule de Bayes-Laplace. Si nous nous posons le problème de savoir quelle probabilité nous avons de nous tromper dans une affirmation concernant p quand nous n'émettons de jugement qu'après chacune des épreuves où E s'est réalisé, "alors il semble, écrit M. von Mises, que la méthode des intervalles de confiance, rejetant toute considération de la distribution initiale, ne peut fournir aucune contribution à la réponse de cette question particulière."[12]

Mais il y a certaines situations où cette seconde forme du problème, que M. von Mises a eu raison de bien préciser, n'est pas celle qui est pratiquement à résoudre. Dans les essais industriels, par exemple, il s'agit, pour les opérateurs, d'accepter ou de rejeter tel ou tel lot suivant le résultat de l'épreuve. Il n'est pas alors possible d'affirmer que ces opérateurs ne se trompent jamais, mais il est important qu'ils se trompent le moins souvent possible.

Supposons maintenant qu'un opérateur ait souvent à décider quelle valeur attribuer à p et à chaque fois d'après le résultat d'une seule épreuve. épreuve. Et conseillons lui d'adopter la conclusion à laquelle nous avait conduit le lemme de Cournot, sans pourtant, cette fois, justifier cette méthode par ce lemme. Nous disions dans le "Fondements des méthodes statistiques d'estimation":

"On peut adopter la règle suivante. On décide avant l'épreuve, de se conformer, après l'épreuve à la règle A_ω consistant:

"Si E se produit, à affirmer ensuite que $\omega \leqslant p \leqslant 1$, si E ne se produit pas, à affirmer ensuite que $\omega \leqslant q \leqslant 1$, c'est à-dire $0 \leqslant p \leqslant 1 - \omega$, et dans les deux cas, d'agir en conséquence. (Observons d'ailleurs que l'on ne pourrait avoir $1 - \omega < p < \omega$, que si l'on suppose $\omega > \frac{1}{2}$, ce qui, du reste, n'a jamais lieu pratiquement, puisqu'on suppose qu'en écrivant $p < \omega$, on n'a pour p qu'une petite probabilité.)

"Ceci étant, quelle est la probabilité P pour qu'on commette une erreur en appliquant cette règle? Si $p < \omega$, on aura commis une erreur si E s'est réalisé: $P = p < \omega$ si $q < \omega$, on aura commis une erreur si E ne s'est pas réalisé: $P = q < \omega$. Enfin si $\omega \leqslant p \leqslant 1 - \omega$, on ne commettra jamais d'erreur, que E se soit réalisé ou non; $P = 0 < \omega$.

"Ainsi l'événement G consistant à commettre une erreur en appliquant la règle A_ω ci-dessus est un événement aléatoire qui a une probabilité P, inconnue comme p, mais tout aussi bien déterminée p et qui de plus est nécessairement inférieure à ω.

---

[12] "On the correct use of Bayes' formula," Annals of Mathematical Statistics, Vol. XIII (1942) p. 165.

"La probabilité P qu'on a de faire une erreur en appliquant la règle A_ω après l'épreuve est déterminée *avant l'epreuve* et elle est $< \omega$. Faisons maintenant un grand nombre d'épreuves indépendantes, dans chacune desquelles l'événement E a une probabilité déterminée mais inconnue, elle sera déterminée avant chaque épreuve et toujours inférieure à ω.

"Il en résulte que si l'on un grand nombre d'épreuves, et si, par exemple, $\omega = 1/10^6$ on devra s'attendre, avant de commencer ces épreuves, à ce que, dans ce grand nombre d'épreuves, on ne fera qu'une erreur au plus sur environ un million de jugements basés sur la règle A_ω. C'est une second manière de motiver les expressions de "coefficient de confiance ou de "probabilité fiduciaire" appliquées à $1 - \omega$.

"Notons bien qu'il s'agit toujours dans ce qui précède de probabilités déterminées avant toute épreuve et sans que la valeur (déterminée) de p soit connue."

C'est là la seule solution (au problème posé dans l'enquête) figurant dans les réponses de MM. Geary, Kendall, Tippett, Darmois, M. Fortet, (ce dernier rappellant que M. Kolmogoroff a publiè un article proposant aussi cette interprétation de confiance). Et M. Neyman précise qu'ils adoptent cette interprétation quand il s'agit en effet de donner une règle d'action—et de la motiver. M. Bartlett distingue: "L'utilité de cette solution dans chaque problème sera une question d'opinion. Il est possible d'imaginer un type de problème, par exemple dans une suite d'essais industriels, ou de telles règles soient tout à fait justifiées et un autre type, peut-être d'un caractère unique, ou la notion d'une estimation regardée seulement comme l'une d'une longue serie d'estimations, peut être plutôt absurde."

En tous cas, en se plaçant dans le premier cas envisagé par M. Bartlett, l'objection faite très justement et rappelée plus haut a une signification du coefficient de confinance $1-\omega$ comme une probabilité ou une chance calculée quand l'épreuve est faite et relative à une valeur de p déterminée avant l'essai, n'a plus de valeur car l'interprétation de $1-\omega$ est celle d'une probabilité d'un événement non encore réalise: west la plus grande valeur de la probabilité qu'on fera (au futur) une erreur en appliquant la règle A_ω. "Sans doute, dit M. Darmois, la probabilité dont nous parlons est relative à la règle adoptée et, quand on l'a appliquée après l'épreuve, on ne sait pas quelle chance on a d'avoir raison. Il est même difficile de donner un sens précise à cette question..."

Il semble donc qu'en adoptant cette interprétation de l'intervalle de confiance, on pourrait mettre d'accord partisans et adversaires de l'emploi de la formule de Bayes-Laplace. Il resterait pourtant à déterminer quelle est la véritable position prise par M. R. A. Fisher.

Par exemple, dans une de ses conférences, M. Fisher s'exprime ainsi: "Cependant le point de vue par lequel je diffère de ceux de quelques autres statisticiens est que nous ne pouvons pas ignorer que nous avons une connaissance certaine de G. C'est la raison même pour laquelle nous avons pris notre échantillon." En replaçant cette phrase dans son teste, il semble bien en résulter que M. Fisher repousse, pour la probabilité fiduciaire, la planche de salut constituée par l'intervention d'une règle d'action applicable à tous les cas possibles.

Dès lors, il semblerait que les critiques formulées à l'encontre de la probabilité fiduciaire par d'aussi éminents statisticiens que MM. Serge Bernstein, Gini et von Mises, si elles ne s'appliquent pas à l'interprétation par la règle A_ω pourraient s'appliquer au point de vue de M. Fisher et en tout cas s'appliquent certainement aux exposés de plusieurs de ses disciples.

C'est ainsi que, selon M. Gini: "M. R. A. Fisher, dans un mémoire de 1930, regarde comme évident—car il ne le prouve pas que s'il y a une probabilité de 0.95 (ou 0.997) par exemple, que la valeur observée, t, d'une grandeur ... d'intensité donnée $\bar{\theta}$ ... excéde la limite F, alors il y a aussi la probabilité 0.95 (ou 0.997) que, ayant observé une valeur E, la valeur de la grandeur reste inferieure à $\bar{\theta}$. Fisher ne fait pas d'autre hypothèse que la continuité de t. Mais la pensée de M. Fisher est fluide et variable. D'une part, il semble avoir depuis ajouté une autre hypothèse à celle de la continuité: celle que la fonction envisagée est ce qu'il appelle "efficient." D'autre part, dans une lettre privée où je lui signalais qu'un de ses disciples avait adopté l'équivalence entre probabilités distinctes signalées ci-dessus par M. Gini, M. Fisher avait refusé d'en assumer la responsabilité et d'accepter cette équivalence. Tout cela n'aurait pas d'importance si nous ne nous trouvions avoir affaire ici à un statisticien dont la réputation est légitimement mondiale et dont nous ne pouvons négliger les écrits. Or nous arrivons à ce dilemme: Ou bien l'interprétation due à M. Fisher de la probabilité fiduciaire est inacceptable, ou bien dans le cas où cette interprétation serait valable, les plus grands statisticiens l'auraient comprise inexactement et il faudrait en conclure que cette interprétation était obscure. Il serait donc urgent d'en donner, dans les ouvrages de statistique adoptant les théories de M. Fisher, un exposé assez clair pour qu'elle ne puisse faire hésiter pas plus les simples usagers de la statistique ceux qui ont apporté à cette science d'importantes contributions. M. E. B. Wilson écrit par ailleurs: "Il y a beaucoup de cours sur la statistique et un nouveau livre sur ce sujet apparait sur le marché à peu près tous les deux mois. Quel est le niveau intellectuel de ces cours et de ces livres? Pour la plupart, autant que j'en puisse juger, ce niveau est le plus bas qu'on puisse imaginer... On y présente une quantité de tables, de méthodes de calcul, de tableaux préparés, en négligeant leur analyse logique..., sans une introduction substantielle au calcul des probabilités[13] qui, avec la logique, fournit le seul fondement solide sur lequel peut reposer une saine compréhension de cette science difficile et traitresse qu'est la statistique."

INFLUENCE DE LA CONCEPTION DE LA PROBABILITÉ

Notons que dans le raisonnement relatif à la probabilité d'erreur de la règle A_w, il n'a pas été nécessaire de faire appel, en dehors de la théorie axiomatique des probabilités, à son interprétation dans le monde sensible. Cette interprétation intervient au contraire, au moment où sachant que $P < \omega$, on doit décider s'il est bon d'appliquer la régle A_w. P peut être interprété par exemple, comme le font MM. Finetti et Fortet, comme un degré de croyance, ou suivant la définition classique de Laplace (encore très généralement adoptée), ou suivant M. von Mises comme une limite de fréquences. M. Neyman adopte une définition qui a en commun avec celle de M. von Mises l'utilisation des fréquences, mais ne suppose pas l'existence nécessaire de leur limite; pour lui, la valeur de la probabilité doit être interprétée comme "la fréquence idéalisée de l'événement dans de longues séries d'épreuves répétées." C'est une définition qui nous semble se rapprocher de celle préconisée par M. Gini et qui parait s'accorder plus encore avec celle que nous avons proposée et où la probabilité d'un événement E dans une catégorie d'épreuves (ou une "population") C, est une grandeur physique attachée au système (E,C), grandeur dont une évaluation expérimentale est fournie par la fréquence de E dans un grand nombre d'épreuves de C.

On trouvera une discussion approfondie de ces différentes définitions dans notre récent article "Les définitions courantes de la probabilité."[14] Nous y avons en particulier répondu d'avance à une objection faite ans sa réponse par M. von Mises contre la définition de Laplace "Il n'y a rien, dit-il, qui puisse être considéré comme des cas également vraisemblables quand un dé irrégulier est jeté. Par conséquent cette interpretation de Laplace doit être rejetée." Dans notre article cité, nous faisons observer à la page 155, que la même difficulté se présenterait dans la théorie des aires si l'on se contentait de définir l'aire par décomposition en carrés égaux. On complètera donc la définition de Laplace comme on complète la définition précédente de l'aire, on conviendra que: "Quand un événement E entraine un événement F, la probabilité de

---

[13] C'est dans le même esprit que M. Wilson, que nous avons consacré notre Premier Cahier de "Leçons de Statistique Mathématique" à un exposé abrégé de calcul des probabilités, exposé dont un second tirage, sous forme polycopiée vient de paraître.

[14] *Revue Philosophique*, 1946, T. CXXXVI, pp. 129-169.

F est par définition au moins égale à celle de *E*." Nous préférons cependant l'interprétation de la probabilité par les fréquences et nous avons répondus (pp. 149-154 de notre article) à certaines objections qu'elle a rencontrées.

Enfin, en ce qui concerne le degré de croyance, nous avons posé cette question: comment pourrais-je me rendre compte si le degré de croyance que j'attribue à un événement se révèle un bon guide de mes actions? Comme il sera certainement un mauvais guide dans quelques cas—l'événement isolé défiant les prédictions,—je n'aurai d'autre resource que de m'assurer s'il y a eu dans le passé une grande fréquence de prédictions justes, c'est-à-dire que j'aurai recours à la notion de fréquence que je voulais éviter. D'autre part' d'après M. Neyman, "l'interpretation de la valeur de la probabilité comme une measure de 'l'intensité de la croyance' appliquée aux événements actuels est à la fois mystique et nuageuse. Si quelqu'un dit que la probabilité que la vie existe sur la planète Mars est égale à, disons 0.357 et que ce nombre represente l'intensité de sa croyance, je puis seulement dire que je ne le comprends pas."

Mais ce n'est pas ici le lieu d'entrer dans les détails de la discussion. Il nous suffit d'avoir rappelé combien les mathématiciens et statisticiens divergent dans leurs interprétations de la probabilité. Or, dit M. von Mises: "Difficultés, malentendus, et confusions sur le problème de l'estimation se sont présentées et doivent nécessairement se présenter tant que le terme probabilité est employé d'une façon vague et n'est pas fondé sur une définition précise."

De son côté, M Fortet écrit: "La théorie de l'estimation des paramètres, et plus généralement de la recherche des causes et des tests d'hypothèses, est essentiellement conditonnée par l'opinion que l'on a sur la nature de la probabilité et sur ce qu'il faut entendre par 'élément aléatoire." D'ailleurs, ces deux auteurs n'adoptent pas la même définition de la probabilité.

On voit ici, se renouveler sur un domaine tout à fait distinct des précédents, les divergences que nousavons signalées au début de ce rapport et qui montrent combien la question enlitige est épineuse.

Cependant, nous plaçant sur le terrain des faits, constatons d'une part, que la technique exige qu'on adopte une solution, bonne ou mauvaise, du problème de l'estimation statistique des paramètres, d'autre part que les résultats fournis par les diverses méthodes envisagées, même quand ils ne coincident pas exactement, sont généralement acceptables dans la pratique. Bien souvent dans de nombreux domines, des règles empiriques (comme l'emploi des coéfficients de sécurité des ingénieurs) ont permis de résoudre les problèmes pratiques posés, sans avoir une justification suffisante. L'adoption dans ces domaines, d'une nouvelle règle, mieux justifiée, quand elle a eu lieu, n'a pas eu pour effet de resoudre un problème déjà pratiquement résolu, mais seulement d'en améliorer la solution.

Le but modeste mais cependant extrêmement utile de cette enquête nous parait être de donner une interprétation plus exacte des méthodes d'estimation des paramètres et nous espérons que ce rapport aura fait faire un pas de plus vers ce but.

Annexe

RÉPONSES A L'ENQUETE

Question posée. Un événement E a été observé r fois dans n épreuves. Sachant (ou admettant) que E y avait une probabilité constante (mais inconnue) p, que peut on dire de la valeur inconnue de p, connaissant seulement n et le nombre r de répétitions de E (ou sa fréquence $f = r/n$?)

A-t-on des objections (et lesquelles?) à faire contre l'une ou l'autre des méthodes connues qui donnent une réponse à cette question?

Pour simplifier à la fois la réponse et la comparaison des réponses, on ne demande d'examiner que le cas où $n = 1$ et $r = 1$.

Réponses.

1. *R. von Mises* (Boston, U.S.A.)
2. *R. Fortet* (Caen, France)
3. *Bruno de Finetti* (Trieste)
4. *Henri Eyraud* (Lyon, France)
5. *Maurice Dumas* (Paris, France)
6. *M. S. Bartlett* (Cambridge, England)
7. *J. Neyman* (Berkeley, California, U.S.A.)
8. *Georges Darmois* (Paris, France)
9. *L. H. C. Tippett* (Manchester, England)
10. *Maurice Kendall* (London, England)
11. *R. C. Geary* (Dublin, Ireland)
12. *Lucien Féraud* (Geneva, Suisse)
13. *Oskar Anderson* (Kiel, Germany)
14. *J. B. D. Derksen* (The Hague, Netherlands; at present Statistical Office, United Nations)
15. *Charles Jordan* (Budapest, Hungary)
16. *Harald Cramér* (Stockholm, Sweden)

1. R. von Mises

THE ESTIMATION PROBLEM

It is most gratifying that the International Statistical Institute decided to initiate a discussion of the estimation problem and thus to bring into

the open an issue that has become highly controversial during the past decades. Difficulties, misunderstandings, and equivocations of the kind here encountered are bound to arise as long as the term Probability is used in a loose way and not based on a precise definition. On the other hand, all the confusion disappears if a clear probablity concept is applied. As a matter of fact, it was this type of difficulties as they arise in connection with the Bayes problem, estimation, small samples, etc., that prompted the present writer to develop, about twenty-five years ago, the fundamentals of a consistent system of probability calculus, generally known today as the frequency theory of probability.

It is not feasible to start the comment on the estimation problem with a statement that these comments are or are not based on "Bayes formula." First of all, just what is meant by "Bayes formula" is to a certain extent controversial. Besides, the strictly negative answer to the particular question ($n = 1$, $r = 1$) submitted for discussion by the Institute has certainly nothing to do with Bayes formula or any formula at all.

In the following a brief discussion of the problem is given. A summary of the results is added at the end.

1. *The basic problem*

The problem proposed for discussion by M. Fréchet has to be restated in the following way if the term Probability is replaced by its definition.

Given the description of a repeatable experiment that can lead to two different results, only, say, zero and one (the result one may also be called occurrence of the "event"). It is supposed that the experiment can be repeated endlessly and that in this sequence the frequency of events approaches a certain limiting value p (which is called the event probability).[1] Now, it is known that n experiments have been carried out and that r times the event has occurred. The question is: What can be inferred from the data n and r upon the value of p, that is, upon the frequency of the event in the long run? In particular, the answer is requested in the case $n = 1$, $r = 1$.

Were the so-called classical definition of probability used, the question would read: What can be inferred from the knowledge of n and r upon the ratio of the number of favorable cases to the number of all equally likely cases? But, there is no such thing as equally likely cases when a biased die is tossed, etc. This interpretation must, therefore, be discarded.

---

[1] In the present text the terms Probability and Chance are used indistinctly for the frequency limit in an infinite sequence of trials. For a more elaborate study of the problem it is advisable to restrict the use of one term (Probability) to those cases where the randomness condition is fulfilled.

On the other hand, the more elaborate scheme of frequency theory that is known today under the heading Probability as a Measure of Set, could be applied. The conclusions, however, would be essentially the same as those presented below.

2. *Answer in the case* n = 1, r = 1.

The answer is simple, short, and unequivocal in the case $n = 1$, $r = 1$, that is when one experiment has been performed and this has led to the event. The answer reads: *nothing.* No inference can be made from the outcome of one experiment or, as will be seen later, from any small number of experiments.

Let, *e.g.*, the experiment consist of tossing a die and observing whether or not six spot appears. If it is known that in one trial the event occurred, it can be that the die is unbiased, *i.e.*, $p = 1/6$; it can also be that that three of the six faces of the die are marked "6"; it can be that the die is loaded in some way so as to make the six spot appear less or more frequently in the long run than with $p = 1/6$; it can even be that six will appear in the long run so rarely that the limiting frequency is zero. No icidation whatsoever is given by the outcome of the single experiment as to which one of these cases (or which other one) is actually present. The same situation prevails if r and n are any small numbers.

If the experiment consists of extracting seven times a letter from a box containing the twenty-five letters of the French alphabet with the appearance of the word MIRACLE as the event, the same conclusions hold. If in one trial the event actually occurred, it can be that the method used in extracting the letters favors some combinations and it can also be that no such bias is present, that is, that in a sufficiently long sequence of trials each combination of 7 letters would appear with the same frequency $p = 4\times10^{-10}$. In the latter case the method of extraction would be said to be unbiased. But, whether in the case which we call "unbiased" the event probability is 1/6 or 10^{-10} is entirely immaterial. It makes no sense to say that an event whose probability is 10^{-10} is "practically impossible." If his way of talking is consistently used one would have to say that no combination of 7 letters is "practically possible." The meaning of $p = 10^{-10}$ is nothing more and nothing less than: In a sufficiently extended sequence the event will appear with this frequency; no additional or modifying statement is justifiable.

The description of the experiment can also be modified so as to read: Extract 7 letters and look up whether or not the outcome appears among the entries in the Dictionnaire de l'Academie Francaise. Here, the p-value in the case which we call "unbiased" will be somewhere between 4×10^{-10} and 1/6. The conclusions remain unchanged.

3. *The nature of the failure*

Why is the answer to the above question totally in the negative? Why is it that no inference can be made from $n = 1$, $r = 1$ upon the unknown value of p? Answer: Because the data are incomplete.

If a surveyor has measured the length of a basis AB and the angle α between the basis and the straight line AC towards some point C, he can make no inference upon the distance AC. Everybody familiar with Euclidian geometry knows that these data are incomplete and that, *e.g.*, the angle β between the basis and BC need also be known.

It will be shown in the following that, in the case of the estimation problem, a certain kind of inference can be made if n is a large number. To give here a geometric analogy, consider this problem: A convex polygon of n sides is inscribed to a circle of area 1; how large is the area of the polygon? It is well known that no inference upon this area can be made (except that it lies in certain limits, as p lies in the interval 0 to 1), since the data are incomplete and something more, *e.g.*, $(n-1)$ angles of the polygon, have to be given. However, if n is a very large number one will be inclined to say that the area is approximately 1 or that it approaches unity when n increases indefinitely. In fact, this is correct if an assumption of qualitative character is introduced, for instance, the assumption that while n increases the ratio of the greatest to the smallest sidelength remains bounded (the bound need not be known). Thus the situation is such, that for small n not even an approximate answer is possible without additional numerical data, while for large n an asymptotic approximation can be found provided some assumption about the circumstances of increasing n is used.

This geometric example shows how we have to proceed if we aim at an (approximate) answer to the estimation problem for large n. The first step must be to complete the data so that they allow a definite answer for any n, large or small. Secondly, we have to introduce a plausible assumption concerning the conditions under which n increases towards infinity. Third, it has to be shown that with increasing n the additional data become less and less influential so that the quantity in question reaches a definite limit for $n = \infty$. The result is, of course, valid only insofar as the assumption used is accepted.

4. *Complete data*

If, in the case of a die, we ask for an estimate of the event probability p, we obviously suppose that there exists a multitude of dice with different p-values and that one of those dice has been chosen for the experiments. Furthermore, asking for an estimate of p does not mean that we think to find the actual p-value of the chosen die, since it is likewise obvious that any couple, r, n, however large n may be, can occasionally be obtained with a die of any value p. All estimation theories agree in this respect. They try to make statements only about the chance or probability that the p-value of the die used in the experiments falls in certain limits.

If we substitute here, as it was done before, for the term Chance its definition, the experiments in the sequence to be considered for the definition include each two steps. The first step consists of the selection of a die from a given multitude of dice, the second of casting this die n times. In the latter sequence we have the well known function

$$q_n(p) = \binom{n}{r} p^r (1-p)^{n-r} \tag{1}$$

for the probability of obtaining r events in n trials, p being the event probability in each of them. If then, in the sequence of first steps, the chance of selecting a die with a p-value between p_1 and p_2 is given in the form

$$\int_{p_1}^{p_2} f(p)dp \tag{2}$$

we have all data required for computing the probability of p as inferred from he n trials. Quite elementary and incontestable rules of probability calculus supply for the probability of p falling in the interval from p_1 to p_2 the expression

$$p_n(p_1, p_2) = \frac{\int_{p_1}^{p_2} f(p)q_n(p)dp}{\int_{0}^{1} f(p)q_n(p)dp} \tag{3}$$

This is sometines called Bayes' formula. (In fact, Bayes did not include the decisive factor, f, neither did Laplace). The correctness of (3) under the assumptions made in the first lines of this paragraph is not contested. Those statisticians who decline to use this "Bayes Formula" do it for various reasons: a) because f is unknown which makes the formula useless; or b) since such a function f does not exist or c) because p is "not a variable but an unknown constant" which means that obivously no $f(p)$ can be considered. We shall deal with the objections b) and c) later. But it has been our task so far to find only that complete set of data which would answer the question for any r and

n. We learn that what is needed in addition to r and n is the function $f(p)$, that is, a definite knowledge independent of, or prior to, the tossings and concerning the structure of the set of dice from which the actual die has been chosen.

5. *Case of large* n

In order to obtain an approximation for $p_n\ (p_1, p_2)$ in the case of large n we study the limit of p_n for n tending towards infinity. A limiting function can be found if, with increasing n the ratio r/n approaches a certain value α of the open interval 0 to 1. Making the following assumptions for f: a) $f(p)$ is bounded in the interval 0 to 1; b) f is positive and continuous at the point $p = \alpha$, we can prove

$$\lim_{n=\infty} = \frac{P_n\ (p_1, p_2)}{\phi(u_1, u_2)} = 1 \tag{4}$$

where ϕ is the Gauss integral

$$\phi(u_1, u_2) = \frac{1}{\sqrt{\pi}} \int_{u_1}^{u_2} e^{-u^2}\, du \tag{5}$$

and the u and p are related by

$$\frac{u_1}{p_1 - \alpha} = \frac{u_2}{p_2 - \alpha} = \sqrt{\frac{n}{2\alpha\,(1-\alpha)}}\,. \tag{6}$$

The convergence in (4) is uniform for all u_1 and u_2 in any interval $-$U to U. The unknown function $f(p)$ does not appear in (4) to (6); no value of F need be given.[2]

If the denominator in (4), in the case of large n, is considered an approximation for the numerator we find for example: If an event occurred in half of all experiments, the probability of p falling in the interval 0.49 to 0.51 is about 0.47 in the case $n = 1000$, and about 0.95 in the case $n = 10{,}000$. The degree of approximation (the strength of convergence) depends, of course, on the upper and lower bounds of f and, above all, on the variability of f in the vicinity of $p = \alpha$. In many cases some general knowledge concerning $f(p)$ is available which allows us to derive exact inequalities for $p_n(p_1, p_2)$.[3] But the main result is this: ***While, from a small number of tossings with one die, nothing can be concluded***

---

[2] This theorem was first proved by R. v. Mises, *Mathem. Zeitschrift*, 4 (1919) p. 83. The conditions for $f(p)$ as given in the text can be reduced to weaker ones. See also R. v. Mises, *Monatshefte fur Mathem. und Physik*, 43 (1936), p. 118.

[3] Such cases are discussed by R. v. Mises, *Annals of Math. Statistics*, 13 (1942), pp. 156-165.

without knowing precisely the structure of the set from which this die was chosen, the inference becomes more and more independent of this additional knowledge as n *increases.*

6. *Small sample theories*

In the last few decades some statisticians developed new methods of estimation with the outspoken aim to derive substantial conclusions from a small number of observations. A detailed analysis of these methods cannot be given here, but two significant examples may be quoted.

In a comprehensive survey of Small Sample Theory (1931), Paul R. Rider discusses[4] a sequence of 10 experiments performed on some sleeping drug. The results (observed prolongation of the normal sleep length) lay between −1.6 and 3.7, four were negative, one zero, and five positive, the average was 0.75 and the standard deviation 1.70. From these observations it is concluded, using the so-called Student distribution: The probability is 0.888 that the drug is effective, *i.e.*, that in the long run the average observed value will be positive. The same method applies also to the case $n = 2$ and would then lead, *e.g.*, to the statement: If one observation is 0.63 and another one 0.67 we have, without knowing anything else about the nature of the drug, a probability of 0.90 that the drug is effective, or in other words: out of 1000 drugs of which nothing else is known than that they gave in one couple of trials those results, about 900 will have in the long run a positive average. In the opinion of the present writer nothing at all can be inferred from 2 observations. In fact, Rider's conclusion is based on a misinterpretation of the Student distribution.

In a paper of 1931, R. A. Fisher discussed at length the exploration of a correlation coefficient from the observations on a small sample.[5] If $n = 4$ pairs of observations give a value $r = 0.99$, he concludes, there is a probability of 0.80 that the long run correlation coefficient ρ will be not less than 0.765. Again this conclusion, so contrary to common sense, is based on an unwarranted substitution of one probability for a different one. It would be extremely dangerous if anybody who has to make a responsible decision about the correlation of two variables were to rely in any way on 4 pairs of observations.

As mentioned above, the usual objection against the classical theory of estimation is that the unknown event probability p (or any quantity derived from it) cannot be considered as the independent variable of a frequency distribution $f(p)$. This may be true in many problems, but it is hard to see what in such cases estimation means. If it is not possible to figure that the object of experimentation can be varied so as to admit various p-values, how can one speak of the chance of p falling in some interval? And, if statements are made that include other terms than chance or probability, like likelihood, fiducial probability, level of confidence or of significance, the first thing to do would be to give clear operational definitions for these terms. This has so far not been done.

---

[4] Paul R. Rider, *Annals of Mathem.*, 31 (1930), p. 577.

[5] R. A. Fisher, *Proceedings Cambridge Philos. Society*, 26 (1930), p. 532.

7. *Summary*

If an inference has to be made from the outcome of n experiments upon the value of the unknown event probability (or any quantity derived from it) sufficient additional information about the set from which the object of experimentation has been chosen is indispensable. For small n, the complete frequency distribution $f(p)$ in this set must necessarily be known. As n increases the influence of this additional knowledge becomes less and less pronounced. For large n, approximative statements are possible, based, as to f, on general, qualitative assumptions only. "Small Sample Theories" which lead to substantial inference for all n, down to $n = 2$ or 3, etc., without using any information prior to the experiments are contrary to common sense and based on a misinterpretation of formulæ of probability calculus.

Résumé

Pour faire une conclusion sur la probabilité p d'un événement qui s'est réalisé r-fois dans n épreuves (ou sur une quantité dérivée de p), il est indispensable de se servir d'une information additionelle suffisante, préalable aux épreuves et concernant la structure de l'ensemble d'objets duquel l'objet employé dans les épreuves fut choisi. Dans le cas de n petit il faut connaître complètement la fonction de distribution $f(p)$ dans le dit ensemble. Lorsque n augmente, l'influence de cette information additionnelle devient de moins en moins prononcée. Pour n suffisamment grand, on peut rédiger des énoncés de nature approximative, basés, quant à f, sur de seules hypothèses générales et qualitatives. Les "Théories d'épreuves en petit nombre"(*Small Sample Theories*) qui fournissent des conclusions substantielles dans le cas de tout n, même $n = 2$, 3, etc., sans avoir recours à une information préalable aux épreuves, sont contre le bon sens et se fondent sur une interprétation fautive des formules du calcul des probabilités.

2. R. Fortet

REPONSE BASEE SUR LA FORMULE DE BAYES

1. *Bases d'une méthode rationnellement satisfaisante*

La théorie de l'estimation des paramètres, et plus généralement de la recherche des causes et des tests d'hypothèses, est essentiellement conditionée par l'opinion que l'on a sur la nature de la probabilité et sur ce qu'il faut entendre par élément aléatoire."

Probabilité et éléments aléatoires: En considérant que tout élément aléatoire peut être ramené à un événement aléatoire, je simplifierai le langage en ne parlant que d'événements.

Je me range (actuellement) à peu près à l'opinion de Keynes. considérant que la probabilité d'un événement est une résultante du corps de renseignements supposé possédé relativement à cet événement: cette façon de voir me parait la plus proche du processus mental de "l'homme de la rue" lorsqu'il conjecture; elle échappe à des objections classiques si l'on remarque qu'on peut inclure dans le corps de reseignements des résultats d'observations empiriques; s'il était prouvé (mais je ne crois pas ce point bien éclairci) que d'un corps de renseignements donnés résulte une probabilité bien déterminée, toujours la même quel que soit l'observateur qui raisonne (sous réserve qu'il raisonne juste), cette probabilité n'aurait rien de subjectif, elle serait purement rationnelle; elle ne serait pas non plus objective dans le sens des théories qui partent de la notion de fréquence empirique.

Il est important de noter que *cette probabilité n'est pas forcément exprimable par un nombre*, il me semble qu'elle n'est assujettie dans le plus général qu'à d'éventuelles relations d'ordre ($>$ ou $<$), et il y a donc une axiomatique à dégager, plus large que celle de Kolmogoroff qu'elle doit inclure. G. Deschamps a fait dans cet ordre d'idées quelques pas intéressants.[6]

Un élément aléatoire n'a donc pas forcément une probabilité numériquement précisée, et *les objections contre la méthode de Bayes fondées sur le fait que dans bien des cas il est impossible de préciser numériquement la loi* a priori *tombent*, sous réserve de formuler la méthode de Bayes dans le cadre la nouvelle axiomatique; je montrerai plus loin comment on pourrait opérer.

Sur une objection classique: On dit souvent (je prends ici le problème précis qui est l'objet de l'enquête) que p n'est pas une variable aléatoire, mais une constante bien déterminée objectivement, simplement inconnue de l'observateur; cette façon a de voir me parait insoutenable; si j'envisage aujourd'hui 28 Janvier 1947, le résultat du 13ième coup de roulette qui se jouera le 15 Juillet 1947 à Monte-Carlo, on sera bien d'accord pour le considérer comme aléatoire, s'agisant d'un coup non encore joué; si je pense à ce même résultat le 20 Juillet 1947, sans avoir eu entre le 15 et le 20 aucun renseignement sur ce résultat, il n'y a rien de change pour moi: cependent on me demanderait à ce moment de modifier mon point de vue, parce que, le coup ayant été déjà joué, son résultat n'est plus aléatoire, mais au contraire bien déterminé, seulement inconnu de moi.

---

[6] Son travail n'est pas publié.

Il est clair que, du point de vue Keynes adopte, p doit être consideréré comme aléatoire.

Extension de la méthode de Bayes: Si p a une densité *a priori* bien définie $f(p)$, sa densité *a posteriori* $\varphi(p)$ est:

$$\varphi(p) = \frac{f(p)p}{\int_0^1 f(u)u\,du} \tag{1}$$

Mais que peut-on faire si p n'a pas une densité *a priori* numériquement précisable ? Je crois pouvoir montrer comment on pourrait procéder de la façon suivante:

Supposons que $F(u) =$ Prob. *a priori* que $p < u$ n'ait pas de valeur numériquement précisable, mais qu'on puisse assigner 2 nombres positifs l et L $(0 < l < L)$ tels que:

$$l(u' - u) \leqslant F(u') - F(u) \leqslant L(u'-u) \qquad u' \geqslant u$$

ce qui traduit des propriétés qualitatives assez simples de la loi *a priori* de p; on peut à peu près traduire cela en disant que $f(p) = \frac{dF(p)}{dp}$ est imprécisée dans la classe C des densités telles que: $l \leqslant f(p) \leqslant L$.

Soit alors C' la classe des fonctions $\varphi(p)$ définies par (1) lorsque $f(p)$ varie dans C: *on pourra considérer comme vraies, pour la loi* a posteriori *de* p, *celles des propriétés de* φ(p), *qui sont communes à toutes les fonctions de la classe* C'.

Pour ne pas allonger cette réponse, je ne pousse pas l'idée précédente jusqu'à ses conséquences numériques, je pense que ce qui précède suffit à le définir.

Evidemment, pour être complètement mise sur pied, cette méthode exigerait quelques travaux, les uns philosophiques, d'autres sur l'axiomatique, etc. Mais pour le moment, je n'ai d'espoir qu'en elle pour résoudre les difficultés de principe que soulèvent les probabilités des causes.

2. *Remarques sur les autres méthodes couramment utilisées*

Les méthodes qui prétendent ne pas s'appuyer sur Bayes sont obligées de se fonder sur un principe plus ou moins arbitraire, qui constitue un postulat nouveau: je critique l'introduction, à mon sens inutile, d'un nouveau postulat.

Mais on peut montrer que, toutefois sous des conditons, mais très générales, *ces postulats peuvent être justifiés par l'application de la méthode de Bayes,* mais seulement asymptotiquement, c'est-à-dire pour n grand; dans le cas de petits échantillons, la précision numérique de ces méthodes est certainement assez illusoire.

A propos des probabilités fiduciaires: M. von Mises a fait justement observer que, étant donnée une règle d'estimation, ce qui importe est de connaître la probabilité d'erreur qu'elle comporte *une fois l'expérience faite*, c'est-à-dire *a posteriori*, et l'évaluation d'une telle probabilité d'erreur *a posteriori* exige, en principe, le recours à la méthode de Bayes.

Toutefois, il faut tenir compte d'une remarque de Kolmogoroff distinguant deux cas:

a) Si celui qui définit la règle d'estimation doit opérer lui-même, et n'est restreint par aucune restriction de temps ou d'argetn, etc., il désirera tirer le maximum de ses résultats, le point de vue von Mises devra jouer;

b) Si celui qui définit la règle d'estimation la définit à l'usage d'exécutants auxquels on ne peut demander de faire des calculs compliqués (cas de celui qui rédige un règlement d'artillerie où il doit définir une règle pour le réglage des tirs), il est logique qu'il envisage seulement la probabilité d'erreur *a priori*; or celle-ci, si elle exige en principe l'intervention de la loi *a priori*, peut souvent être évaluée sans sa connaissance:[7] le point de vue fiduciaire prend alors, et en toute rigueur, tout son intérêt.

3. B. de Finetti

I. *REPONSE BASEE SUR LA FORMULE DE BAYES*

(1) En effet le problème et les méthodes n'ont de sens qu'en s'appuyant sur le raisonnement de Bayes, même s'il est sous entendu, déguisé, refusé;

(2) chercher une justification "sans recours à la formule de Bayes" c'est comme vouloir conclure de quel côté penchera une balance du fait que l'un des poids est "petit" en grandeur absolue (4) (au lieu que "petit" par rapport à l'autre (5));

(3) il n'y a rien de frappant si un événement très peu probable (6) (que ce soit le mot MIRACLE ou le mot CDATIPS) se réalise, et l'on ne peut donc du fait de sa réalisation conclure que l'événement "avait une probabilité moins petite"(10);

---

[7] Voir l'exemple envisagé par M. Fréchet dans sa note à Portugaliae Mathematica, 1946, pp. 137-141.

(4) de même il n'y a aucune raison, de ce fait, de soupçonner que le tirage ayant donné résultat quelconque (MIRACLE ou CDATIPS) était truqué (ou, en général, influencée par une cause A déterminée);

(5) le soupçon de trucage ou l'intervention de la dit cause A est justifié non-de ce fait que le résultat vérifié était "très peu probable" (fait identique pour MIRACLE et pour CDATIPS); mais par comparaison avec la probabilité qu'un trucage ait été établi dans le but de réaliser certains résultats (probabilité grande pour MIRACLE et petite pour CDATIPS);

(6) par ailleurs, la "probabilité du résultat vérifié" n'est pas une notion définie (objection de M. Dumas: le "résultat," est-ce le mot "MIRACLE" ou les faits "mot français," "mot d'une langue connue," "combinaison prononçable," etc.?); par conséquence les procédés "sans Bayes" ne sont pas seulement non justifiés mais non susceptibles d'application univoque (au contraire: elle est totalement indéterminée);

(7) cette difficulté s'évanouit en adoptant le raisonnement de Bayes où interviennent les *rapports* entre les probabilités (p. ex. probabilité truquée et probabilité au hasard de "mot français" ou du mot "MIRACLE; le rapport ne change pas si tout mot français a la même probabilité truquée; dans la supposition contraire le rapport pour "MIRACLE" donne l'évaluation correcte qui tient compte aussi des raisons particulières qu'on a de s'attendre un trucage destiné a produire précisement ce mot).

II. *Jusqu'ici je crois que le point de vue exposé coincide avec le point de vue classique; la conception "subjective" qui répond à la façon de penser qui m'est propre ne modifie nullement ces conclusions mais les rend beaucoup plus nettes en adoptant les prémisses plus radicales suivantes*:

(8) la probabilité d'un événement est la mesure du degré de confiance d'un sujet donné à l'égard de son arrivée;

(9) elle n'est donc définie qu'avec référence à l'opinion d'un individu qui l'évalue (ou à l'opinion moyenne d'un group d'individus) comme la *valeur* d'une marchandise, qui serait une notion dépourvue de sens en considérant des objets dont personne connait l'existence;

(10) l'évaluation d'une probabilité ne constitue pas une *fait* dont on peut dire s'il est "vrai" ou "faux"' bien une opinion;

(11) la notion de "probabilité *inconnue*" n'a pas de signification, la probabilité n'existant qu'en tant que et dès l'instant où elle a été évaluée par quelqu'un;

(12) par conséquent on ne peut pasx parter de la probabilité que la probabilité d'un événement E ait des valeurs données (probabilités de 2^{e} ordre, etc.);

(13) l'évaluation d'une probabilité concerne toujours une épreuve bien déterminée, considerée isolément; que plusieurs épreuves soient appelées "épreuves d'un même événement" ou non ce n'est qu'une commodité de langage dénuée de signification intrinsèque et d'intérêt pour la théorie des probabilités (je préfère la terminologie: événement = épreuve; épreuves d'un même phénomène au lieu que "d'un même événement").

III. *En voici les conséquences pour la position de problème considéré:*

(14) on n'y peut pas parler de probabilité "inconnue," et on ne peut interpréter l'influence du résultat des épreuves successives observées comme une correction progressive dans l'évaluation de la même probabilité inconnue;

(15) au contraire, la probabilité non conditionnelle de chaque épreuve et les probabilités conditionnelles (subordonnées aux divers résultats des épreuves précédentes) sont autant de grandeurs distinctes évaluées (ou de moins évaluables) avant toute expérience; l'effet de l'observation, c'est non de corriger l'évaluation d'une même grandeur mais de choisir parmi cet ensemble de grandeurs données d'avance celle qui correspond à l'état de connaissance atteint;

(16) les difficultés provenant de la distinction entre "probabilité" et "vraisemblance" disparaissent, car selon (8), (13) il s'agit toujours de "probabilités" entendues dans le sens qui pour ceux qui feraient la dite distinction, corresponde plutôt au mot "vraisemblance";

(17) après avoir ainsi éliminées les difficultés que l'on rencontre d'après les autres conceptions et qui, selon la présente, apparaissent illusoires, on recontre par contre des difficultés nouvelles dans la tâche de poser le problème en se passant des notions déclarées illusoires et en gardant des exigences particulièrement rigides de rigueur logique qui en découlent (la chose la plus difficile c'est cependant de faire pénétrer l'esprit de la position du problème à ceux qui sont accoutumés à le voir a autrement);

(18) il s'agit avant tout de définir ceux que l'on appellerait "événements indépendants à probabilité constante mais inconnue"; on y a réussi par la définition de "événements équivalents" (tels que, de toute façon que l'on en choisisse un nombre n quelconque, la probabilité que tous le n événements se vérifient a toujours une même valeur ω_n, ne dépendant que de n; il en résulte qu'il en est de la probabilité $\omega_r^{(n)}$ que r parmi eux se vérifient et n-r non; en particulièr la probabilité de chaque événement a toujours la même valeur ω_1);

(19) de cette définition il découle que la loi de probabilité $F_n(x) = \Sigma_{r<nx}\,\omega_r^{(n)}$ de la fréquence r/n, converge vers une loi-limite $F(x)$

lorsque $n \to \infty$, et que toute probabilité peut s'exprimer *comme si* l'on admettait de dire "la probabilité inconnue, etc., est un nombre aléatoire X suivant la loi $F(x)$" (c'est-à-dire: "$F(x)$ est la probabilité que cette probabilité inconnue X soit $<x$"); p. ex.

$$\omega_r^{(n)} = \int \binom{n}{r} x^r(1-x)^{n-r} dF(x);$$

(20) il est légitime d'associer à ce fait une interprétation concrète seulement s'il y a une grandeur X *physiquement definissable* qui *coincide* avec la valeur de la probabilité conditionnelle: p. ex. dans le cas d'une urne de composition inconnue R/N, en supposant que, comme probabilité subordonnée à la composition R/V, on prenne la valeur R/N, la probabilité conditionnelle coincide avec la composition" (pourcentage des boules blanches) qui est une grandeur physique;

(21) il faut cependant toujours distinguer entre la grandeur (fait objectif) et la probabilité (opinion subjective): même dans l'ex. (20), on n'est pas obligé à la dite évaluation: pour des raisons particulières ou même sans raisons déterminées on aurait le droit de penser p. ex. que chaque boule blanche à probabilité double des noires et ainsi que la probabilité correspondant à la composition R/N serait non plus R/N mais $2R/(2R+N-R) = 2R/(N+R)$;

(22) la position de problème de (15) n'est pas liée a l'hypothèse particulière qui caractérise les "événements équivalents" mais sa validité été tout à fait générale; une condition plus large de l'"équivalence" est, p. ex., celle de l'"équivalence partielle."[8]

IV. *EXEMPLES pour éclaircir quelques-unes de ces thèses*:

(3), (6) Si les raisonnements "sans Bayes" à propos du tirage "MIRACLE" devaient être acceptés, il faudrait accepter de même les sophismes de ces joueurs qui refusent un billet de loterie portant un numéro "spécial" (p. ex. 100,000 ou 123,456) car de l'hypothèse que tout billet ait la même probabilité ils déduisent contradictoirement que celui-là a une probabilité *particulièrement* petite.

(15) Au même compte que la probabilité de gagner une loterie après que le billet donné a été tiré, après qu'il a été associé à un des concourants au grand-prix, après avoir assisté à la course qui va se conclure sous peu, ne sont pas des *corrections* successives de l'évaluation initiale que ait été "incorrectement" fondée sur le nombre des billets achetés, seul élément du jugement dont on pouvait disposer à cet instant.

---

[8] Pour toutes ces questions voir: *La prévision: ses lois logiques, ses sources subjectives* (Inst. Henri Poincaré), *Sur la condition d'"équivalence partielle"* (Actualités scientifiques et industrielles, 739).

(20) Exemple du cas contraire: le jeu de pile ou face avec une pièce d'apparence irrégulière. La loi de probabilité $F(x)$ de la fréquence sur un "nombre très grand d'épreuves" n'est pas la loi de probabilité d'une grandeur physiquement définissable coincidant (dans le sens de (20)) avec la probabilité. On n'a pas le droit de dire que $F(x'')-F(x')$ est la probabilité que "l'imperfection de la monnaie soit telle qu'elle lui donne probabilité comprise entre x' et x'', car ne sont pas définies les circonstances de fait qui correspondraient à telle évaluation de probabilité.

4. H. Eyraud

REPONSE BASEE SUR LA FORMULE DE BAYES[9]

La solution dépend d'une fonction arbitraire, notamment la probabilité *a priori* $u(x)$ pour que la probabilité p inconnue soit comprise entre zéro et x.

Si $U(x)$ est alors la fonction cumulative de probabilité *a posteriori*, c'est-à-dire après l'apparition de l'événement E dans une seule épreuve, on a:

$$U(x) = \frac{\int_0^x x \, du(x)}{\int_0^1 x \, du(x)}$$

Dans le cas le plus simple de probabilité *a priori* uniforme (absence de dominante) on a:

$$u(x) = x \qquad 0 < x < 1$$

$$U(x) = x^2 \qquad 0 < x < 1$$

La probabilité dominante est $p = 1$

La probabilité moyenne est $p = 2/3 = 0.666$

L'écart typique est $\sigma = \frac{1}{3\sqrt{2}} = 0.235$

5. M. Dumas

SANS RECOURS A LA FORMULE DE BAYES

D'après le résultat une fois sur une, il n'y a aucune raison de supposer que l'événement contraire de E, soit E', peut se produire, à moins de faire état d'une connaissance qui ne figure pas dans l'énoncé de la question.

---

[9] *Cf.* H. Eyraud, *Cours de Probabilités et Statistiques*, 2e édition, 1936, p. 17 (Institut de Science Financière et d'Assurances, Université de Lyon).

Le bon sens impose donc de considérer l'arrivée de l'événement E *comme une certitude*, soit $p = 1$; mais il indique par ailleurs que cette certitude n'est basée que sur un très petit nombre d'épreuves et qu'il ne faut peut-être pas lui accorder grande confiance.

Pour avoir une idée cette dernière, je propose ceci: Si une seconde épreuve était faite, elle conduirait au total, soit à deux arrivées de E sur deux épreuves, ce qui renforcerait la certitude; soit à une arrivée de E sur deux épreuves.

Donc, la confiance à accorder à la certitude résultant de 1 fois sur 1, est *au moins égale* à celle qui résulterait de 1 fois sur 2. Or, par raison de symétrie, d'après 1 fois sur 2, le bon sens conduit à attribuer à toute valeur de p une probabilité élémentaire la même dans tous les cas. Donc, d'après le seul résultat 1 fois sur 1, la confiance à accorder à l'estimation $p = 1$, est à préciser ainsi: la probabilité intégrale de toute valeur p_1 de p est *au plus égale* à p.

Remarque I: Extension de la règle A_ω

a) La question posée se présente de façon quelque peu différente si le résultat 1 fois sur une épreuve est constaté au cours d'une série d'ensemble d'épreuves, si par exemple la question est la suivante:

"On dispose de m populations à chacune desquelles correspond une valeur de p qui lui est propre et qui est inconnue. On convient de faire subir une épreuve à chacune de ses populations. On remarque j populations ayant fait apparaître l'événement E et m-j l'événement contraire de E. Que peut-on penser d'une population ayant fait apparaître E?"

D'après la règle A_ω (M. Fréchet, 1946) on a une probabilité de se tromper au plus égale à ω au cours des m affirmation suivantes: populations ayant donné E: pour chacune d'elles la valeur de p est dite comprise entre ω et 1; populations ayant donné événement contraire de E; pour chacune d'elles la valeur de p est dite comprise entre 0 et $1-\omega$.

Cela est certain, mais on peut aller plus loin. On peut affirmer que l'espérance mathématique du nombre des erreurs faites est au plus égale à $m\omega$ définissant m_1 par la condition que d'après la loi binomiale de paramètre ω, et appliquée à m tirages, la probabilité intégrale de $m_1\omega$ est égale à $1-\varepsilon$, on peut affirmer avec la probabilité $1-\varepsilon$ que le nombre total des erreurs faites est au plus égal à $m_1\omega$. On peut donc affirmer *a fortiori* avec la probabilité $1-\varepsilon$ que sur les j populations ayant fait apparaître l'événement E, il y en a *au plus* $m_1\omega$ pour lesquelles l'affirmation que p est compris entre ω et 1 est fausse.

b) La quantité $m_1\omega$ peut être quelconque par rapport à j; il se peu que l'on ait $m_1\omega > j$ et dans ce cas on ne peut même pas affirmer avec la probabilité $1-\varepsilon$ que parmi les j populations ayant donné E, il y en a une seule dont le p soit effectivement compris entre ω et 1.

Voici l'exemple d'un cas où, s'il en était autrement. cela choquerait le bon sens:

Soient m populations dont les p sont tous égaux à $\omega_1 < \omega$. Si ω_1 n'est pas nul, il y aura cependant des populations qui à l'épreuve donneront E; soit j ce nombre. L'espérance mathématique de ce nombre est $m\omega_1$; si l'on définit m_2 par la condition que d'après la loi binomiale de paramètre ω_1, et se rapportant à m épreuves, la probabilité intégrale de $m_2\omega$, *est* égale à $1-\varepsilon$, on sait avec la probabilité $1-\varepsilon$ que le nombre j des épreuves donnant E est inférieur à $m_2\omega_1$. Or d'après les lois binomiales considérées: $m_2\omega_1 < m_1\omega$. Donc $j < m_1\omega$.

c) Voici maintenant l'exemple d'un cas où l'extension ci-dessus de la règle $A\omega$ est d'un grand intérêt pratique: lorsque l'on veut classer des lots d'objets en deux catégories suivant le résultat d'une série d'épreuves subie par chacun d'eux, c'est souvent pour mettre en service les lots d'une catégorie et pour rebuter les autres; et s'il faut éviter de rebuter des lots qui auraient mérité d'être mis en service, l'essentiel est de ne pas mettre en service des lots qui mériteraient d'être rébutes. On sait calculer grâce à l'extension susvisée et en fonction d'une probabilité ε combien, au minimum, parmi les lots admis il y en a qui méritaient effectivement de l'être. L'attention est opportunément attirée sur le fait que ce nombre est fonction du nombre total des lots éprouvés, qu'ils aient ou non été rebutés: cette notion ne nous semble pas avoir été bien dégagée jusqu'à présent, mais maintenant elle nous paraît s'imposer avec la force d'une évidence.

Remarque 2: *Trop grande particularité de la question posée*

La question posée se rapporte à un cas très particulier et il se trouve précisément que le bon sens conduit pour elle à une solution plus complète que dans n'importe quel autre cas; en effet sa solution met en cause le cas de l'arrivée de 1 fois sur 1 qui est un résultat symétrique. Si par exemple la question posée avait été relative à 2 fois sur 2, nous aurions été amené à envisager le cas 2 fois sur 3, cas qui attire l'attention sur la valeur $p = 2/3$, mais qui ne permet pas de conclure relativement aux probabilités intégrales de p.

Je crois devoir rappeler ici un résultat donné par M. Bachelier et qu'admet le bon sens, à savoir que du seul fait que le résultat r fois sur n épreuves a été obtenu, et que d'une façon générale le résultat r_2 fois sur n_2 épreuves attire l'attention sur la valeur $p = r_2/n_2$, on est conduit à donner à l'arrivée de r_1 fois en n_1 nouvelles épreuves la probabilité:

$$\frac{n_1!}{r_1!\,(n_1-r_1)!}\cdot\frac{(n-1)!}{(r-1)!\,(n-r-1)!}\cdot\frac{(r+r_1-1)!\,(n-r+n_1-r_1-1)!}{(n+n_1-1)!}$$

On sait que le raisonnement consiste à admettre que l'expérience nouvelle est suivie épreuve par épreuve et que pour chacune d'elles la probabilité de l'arrivée de l'événement E est égale à la fréquence calculée d'après toutes les epreuves antérieures.

Remarque 3: *Conditions moyennant lesquelles une méthode d'estimation de* p *peut être retenue*

Je pense qu'à moins de faire plus ou moins explicitement une hypothèse, le bon sens, et par suite aussi le calcul, en présence d'un résultat r fois en n épreuves, ne peut pas faire plus qu'attirer l'attention sur la valeur de p $p_1 = r/n$, et qu'en particulier il lui est impossible de préciser quelle confiance on est en droit de reconnaître à cette estimation. Je pense donc que tout raisonnement precisant cette confiance implique nécessairement une hypothèse, et il est de fait que parfois des hypothèses non faites explicitement ont pu être décelées (M. Fréchet; 1943).

Je ne m'oppose pas à ce qu'une hypothèse soit faite pour permettre de préciser cette confiance, mais je pense que l'on peut exiger deux choses: d'une part, *que cette hypothèse soit bien dégagée et explicitement formulée d'autre part, surtout que le développement de cette hypothèse ne conduise à aucune contradiction par rapport à ce qu'indique le bon sens.*

Dans ma réponse basée sur la formule de Bayes, j'indique une hypothèse de probabilité *a priori* qui satisfait à ces deux conditions.

BIBLIOGRAPHIE

Bachelier, *Calcul des probabilités* (Paris, Gauthier Villars), 1912.

Estienne, *Loisirs d'artilleur* (Paris, Berger Levrault), 1906.

Fréchet, "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons," *Revue* Inst. Int. de Stat., 1943.

Fréchet, "Fondements des méthodes statistiques d'estimation," *Portugaliae Mathematica*, 1946.

REPONSE BASEE SUR LA FORMULE DE BAYES

Toute réponse basée sur la formule de Bayes ne peut faire intervenir une loi de probabilité *a priori* déduite d'une connaissance antérieure, puisqu'aucune telle connaissance n'est notée dans l'énoncé de la question. Force est donc de faire une *hypothèse* de probabilités *a priori.* Je propose de retenir comme hypothèse de probabilités *a priori*, la loi virtuelle dont la probabilité élementaire est proportionnelle à $dp/p(1-p)$, ou, plus succinctement, je propose l'hypothèse $dp/p(1-p)$, et cela pour les raisons exposées dans la remarque I.

Avec l'hypothèse retenue, la réponse à la question posée est celle-ci; la probabilité élémentaire d'une valeur p est

$$p \frac{dp}{p(1-p)} \int_0^1 \frac{dp}{1-p}.$$

L'intégrale du dénominateur est infinie; de ce fait à toute valeur de p correspond une densité de probabilité nulle, sauf lorsque p égale 1 car toute cette valeur la densité prend la forme indéterminée $1/0 \times \infty$; comme la probabilité totale des valeurs de p est 1, c'est que cette form; indéterminée est égale à 1. Finalement, la connaissance exclusivede 1 fois sur 1 conduit à conclure à la *certitude* de l'arrivée de l'événement E, comme le veut le bon sens (*cf.* notre réponse sans recours à la formule de Bayes).

Remarquons que le résultat 1 fois sur 2 aurait conduit à

$$p(1-p) \frac{dp}{p(1-p)} : \int_0^1 dp = dp$$

et que par suite, les considérations dévéloppées dans notre réponse sans recours à la formule de Bayes étant encore valables, l'hypothèse $dp/p(1-p)$ conduit *elle aussi* à ceci; d'après le résultat 1 fois sur 1, on doit admettre que la probabilité intégrale de toute valeur p_1 de p est au plus, égale à p_1.

Remarque 1: Raisons en faveur de l'hypothèse retenue

a) L'hypothèse $dp/p(1-p)$ correspond a la loi sur laquelle M. Lhoste (1923) a attiré l'attention en faisant valoir ce qui suit. D'une part, si une variable aleatoire H a pour intervalle de variations $(0, +\infty)$ on ne peut sous peine de contradictions (voir Lhoste 1923 ou Dumas 1947) admettre à titre d'hypothèse une autre loi de probabilité *a priori* que celle caractérisée par dh/h; d'autre part on ramène au précédent le cas de la variable P par le changement de variable $H = (1-p)/P$, car l'intervalle de variations de H est $(0, +\infty)$. Or à dh/h correspond $\frac{dp}{p^2} \cdot \frac{p}{1-p} = \frac{dp}{p(1-p)}$

b) La loi virtuelle $dp/p(1-p)$ *n'est pas mathématiquement absurde.* En effet, pour $\varepsilon < 0.5$ aussi petit qu'on veut, la loi de probabilité

$$\frac{1}{2Lg.\ \frac{1-\varepsilon}{\varepsilon}} \int_{\varepsilon}^{1-\varepsilon} \frac{dp}{p(1-p)}$$

est bien déterminée dans tout l'intervalle de variations $(\varepsilon, 1-\varepsilon)$. L'application de cette loi de probabilité *a priori* au cas général où l'on observe r fois en n épreuves est

$$\frac{1}{2Lg.\frac{1-\varepsilon}{\varepsilon}} C_n^r p^r (1-p)^{n-r} \frac{dp}{p(1-p)} : \int_{\varepsilon}^{1-\varepsilon} \frac{\quad}{\quad} dp.$$

Comme du fait de la division des deux membres par les facteurs indépendants de p, l'expression ne dépend plus de ε que par les limites de l'intégrale du dénominateur, et comme cette intégrale a une valeur bien déterminée pour toutes valeurs possibles de r de n et surtout de ε et même pour la valeur limite $\varepsilon = 0$, il n'est pas mathématiquement absurde de raisonner sur l'hypothèse $dp/p(1-p)$ dans tout l'intervalle (0, 1), limites comprises.

c) L'hypothèse $dp/p(1-p)$ *conduit aux resultats que le bon sens permet à lui seul de dégager.* En voici quelques preuves. Les premières ont été données ci-dessus dans le paragraphe précédent. Une autre est qu'avec cette hypothèse, en présence du résultat r fois sur n, on a bien l'attention attirée sur la valeur $p = r/n$, puisque r/n est *l'espérance mathématique* de la loi de probabilité *a posteriori* à laquelle cette hypothèse conduit:

$$\frac{(n-1)!}{(r-1)!\,(n-r-1)!}\, p^{r-1}(1-p)^{n-r-1}dp.$$

Enfin, avec cette hypothèse, la probabilité de l'arrivée de r fois en n épreuves si l'on sait seulement que dans la même population on a obtenu r fois en n épreuves, est précisément égale à l'expression de Bachelier (voir notre réponse sans recours à la formule de Bayes) à savoir:

$$\frac{n_1}{r_1!\,(n_1-r_1)!}\cdot \frac{(n-1)!}{(r-1)!\,(n-r-1)!}\cdot \frac{(r+r_1-1)!\,(n-r+n_1-r_1-1)!}{(n+n_1-1)!}$$

N. B. Cette expression est différente de celle qu'employait parfois K. Pearson et qui correspond à l'hypothèse dp. D'ailleurs d'une façon générale l'hypothèse $dp/p(1-p)$ *est la seule* qui conduise à tous les résultats rappelés en (a) à (c) ci-dessus; elle est donc *la seule* à pouvoir de ces points de vue être retenue.

Remarque 2: Comparaison de la méthode proposée avec toute autre méthode

a) La méthode proposée est supérieure à toute autre méthode basée sur la formule de Bayes, d'après le N. B. qui précède. Toute méthode sans recours à la formule de Bayes suppose une hypothèse (*cf.* Fréchet

1943) et par suite il nous suffit de montrer en (b) ci-dessous qu'hypothèse pour hypothèse, le mieux est de faire une hypothèse de probabilités *a priori.*

b) Toute méthode basée sur une hypothèse de probabilités *a priori* présente notamment les avantages suivants (voir sur le même sujet mais relativement à une loi de Laplace-Gauss: Dumas 1947): (i) elle permet de déterminer complètement une loi de probabilité *a posteriori* sans qu'aucune approximation n'ait à être faite; (ii) elle permet des développements simples et précis dans le problème traité en (c) de la remarque I, sans que l'hypothèse ait à être faite que la première série donne une idée exacte de la probabilité p inconnue; (iii) elle permet également des dévéloppements simples et précis lorsqu'il s'agit de *comparer* deux résultats, puisqu'un raisonnement non indiqué ici conduit à la loi de probabilité de la variable aléatoire P_1-P_2, difference des deux probabilités inconnues; (iv) elle permet tenir compte de tout ce que l'on connaît, même si cela est vague, et d'en déterminer les répercussions sur la loi de probabilité *a posteriori*: si par exemple on sait r fois sur n et si on sait de plus que p est petit, il peut être convenable de passer par l'approximation de la loi binomiale par la loi de Poisson de paramètre a et de prendre comme hypothèse de probabilités *a priori*, celle caractérisée par da/a.

BIBLIOGRAPHIE

Fréchet, "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons," *Revue* Inst. Int. de Stat., 1942.

Dumas, "Sur une loi de probabilité *a priori* conduisant aux arguments fiduciaires de Fisher," *La Revue Scientifique,* 1947.

Lhoste, *La Revue d'artillerie,* 1923.

6. M. S. Bartlett

WITHOUT RECOURSE TO BAYES' FORMULA

(1) *The confidence interval solution.*

To repeat the form of solution discussed by Fréchet in "Fondements des méthodes statistiques d'estimation," (*Portugaliae Mathematica*, 5: 3 (1946), 137-141):

> We may cover the unknown value p, with a known probability of error not greater than ε by a random interval with realized value $(0, 1-\varepsilon)$ if the event E does not occur in the single trial, and realized value $(\varepsilon, 1)$ if the event does occur. Given that E has occurred, the realized value of the interval is thus $(\varepsilon, 1)$,

(2) *Critical comment.*

The class of estimation problem for which we adopt this form of solution does not affect the validity of the inference (we assume of course the correctness of the theoretical model, including the absence of any selection, depending on the realized data, which would invalidate in any way the particular confidence interval solution made use of); but the usefulness of the solution in any problem will be a matter of opinion. It is possible to imagine one type of problem, *e.g.*, in routine sampling inspection in industry, where such inferences are extremely relevant, and another, perhaps of a unique character, where the notion of an inference regarded merely as one in a whole set of such inferences, may seem rather absurd.

I reject the quantitative use of Bayes' formula in estimation problems because the *a priori* probabilities it requires, whether regarded as subjective or objective quantities, are not in my opinion quantities to which unique and known values may be assigned. But whatever method of estimation is adopted, we should admit that we cannot lay down what any person should *believe* the value of p to be (except the trivial one that he would be inconsistent if he believed the value of p to be smaller after E has occurred than before).

7. J. Neyman

The answer depends on the precise interpretation of the words "probability" and "what can one say."

I. *Interpretation of probability.*

a) According to some authors, the value of the probability is interpreted as a measure of "intensity of belief." To the present author this interpretation, applied to actual events, seems both mystic and misty. If somebody says that the probability of there being life on the planet of Mars is equal to, say, .357 and that this figure represents the intensity of belief, I just do not understand it.

b) According to other authors, and this point of view is shared by the present writer, the value of the probability is interpreted as the idealized relative frequency of the event E in long series of repeated trials. In all that follows, this latter interpretation will be adhered to.

II. *Interpretations of the words: "What can one say."*

a) The words "what can one say" are frequently interpreted as "what is implied." This interpretation is frequent in mathematics. For example, I can think of the following question: "given that a function $f(x)$ is continuous in the closed interval (a, b), what can one say about its integrability in the sense of Riemann?" The answer is: The continuity *implies* the integrability. On the other hand, the integrability does not imply the continuity in the whole interval.

With this interpretation (a) the answer to the question underlying the present enquête is that, the appearance of E a stated number x times in the course of n independent trials implies nothing about its probability p. In fact, the latter may be anything between zero and unity.

Some conclusions about the probability p would be possible if, apart from the data enumerated in the Question under consideration, it were known that the probability p is itself a random variable. In that case the formula of Bayes could be applied and the result of such application would depend on the knowledge of the distribution function of p.

b) "What can one say" may be interpreted as "what should one do" or "how should one behave" if faced with the necessity of some action whose success depends on the value of p. Imagine, for example, that

if $0 \leqslant p < .1$ then our preferred action would be A_1
if $.1 \leqslant p < .2$,, ,, ,, ,, ,, ,, A_2

etc. Whatever our experiments and observations, the actual value of p will remain unknown. However, if faced with the necessity of a choice between a number of possible practical steps $A_1, A_2, \ldots$, we may think of establishing a rule of behavior (inductive behavior) which would make the proposed action $A(x)$ a function of the observed number x of occurrences of the event E in the course of n independent trials. For any such rule, R, we could compute the probability, say $P(R \mid p)$, of the step $A(x)$ chosen being actually the preferred one (this may be interpreted also as the probability of our statement about the value of p being correct). Every value of p implies a value of $P(R \mid p)$. Let $P^*(R) = \alpha$ be the greatest lower bound of $P(R \mid p)$. Since in any actual case the value of p is unknown, we may state only that the probability of the rule R leading to a success is at least equal to α. The number α so defined is called the confidence coefficient relating to the rule R.

I wish to conclude my remarks with a quotation from my paper in the *Journal of the Royal Statistical Society*, Vol. XCVII (1934) pp. 624-625.

"Next, let me remark on the criticism of Professor Bowley. I shall simplify the position, assuming that we want to determine the proportion, p, of black balls in a bag, and for this purpose we intend to draw as many as $n = 3$ balls (with a replacement—for simplicity). The number of black balls in the sample will be denoted by X, which may have the values $0, 1, 2, 3$. Professor Bowley assumes that we actually have the sample drawn and that, for instance $X = 3$. Having got so far, we certainly cannot tell much about the probability of p having any definite value. In fact, all we can say is that either an improbable fact occurred, of p has a rather large value. But what is done in the method of confidence intervals is different. We start our reasoning, as it were, before drawing the sample. Or again, we assume that this is not the only one sample with which we shall have to deal. We notice that the sample may provide us only with one out of the four possibilities $X_1 = 0$, $X_2 = 1$, $X_3 = 2$, $X_4 = 3$. Having noticed this, we fix a rule as follows:

"If in the sample which we shall draw X will have the value

$X = 0$, then we shall state that $0 \leqslant p \leqslant \pi''$
$X = 1$, ,, ,, ,, ,, ,, $\pi'_2 \leqslant p \leqslant \pi_2''$
"If $X = 2$, ,, ,, ,, ,, ,, $\pi'_3 \leqslant p \leqslant \pi_3''$
$X = 3$, ,, ,, ,, ,, ,, $\pi'_4 \leqslant p \leqslant 1$

We are aware that the statement which we shall make, in applying this rule to the result of actual sampling, may be wrong or may be true. We calculate the probability, P, that the statement will be a true one, and try to arrange the system of values of the π's so as to have $p \geqslant .95$, whatever the probability law *a priori*.

"In this way, having a sample and knowing that $X = 3$, we cannot tell any more about the value of p than Professor Bowley has stated. On the other hand, making statements following the rules set out above, we know something important about the results of these statements: the probability that we shall be wrong is then $\leqslant .05$."

8. G. Darmois

SANS RECOURS A LA FORMULE DE BAYES

Que peut-on dire si on a observé une fois l'événement E, de probabilité p (inconnue)?

Ce qu'on peut dire d'abord, c'est que E n'est pas impossible. Et en disant cela, on est assuré de ne pas se tromper. Nous ne serons jamais contredit par un observateur qui, connaissant la valeur exacte de la probabilité p, entend nos réponses et les juge.

Il s'agit de savoir si nous pouvons essayer de donner avec des risques appropriés, des renseignements plus précis. Evidemment, tout réponse précise telle que: la probabilité p est comprise entre $\alpha > 0$ et l'unité, sera ou vraie ou fausse, et dans cette affirmation ou rien n'est aléatoire, il ne saurait être question de probabilité d'erreur.

Il n'interviendra de probabilité que si, considérant l'épreuve comme fournissant un résultat aléatoire, nous convenons d'attacher à ce résultat aléatoire une affirmation, également aléatoire, telle que celle-ci; un certain intervalle, déterminé par l'épreuve, contient la valeur inconnue p.

Dans ces conditions, cela a un sens très déterminé de parler de la probabilité qu'une telle affirmation soit vraie ou fausse.

En particulier, la règle qui consiste à attacher à la réalisation de
l'événement E, l'intervalle $\alpha \leqslant p \leqslant 1$
l'événement non-E, l'intervalle $\alpha \leqslant q \leqslant 1$
possède une probabilité inférieure à α de conduire à une erreur.

C'est-à-dire que l'observateur qui à chaque fois, juge nos réponses faites d'après cette règle, observera une série statistique à laquelle il donnera une série de notes oui ou non, la note non ayant une probabilité inférieure à α.

Tout ceci est très exactement conforme à ce que dit M. Fréchet dans son "Fondements des méthodes statistiques d'estimation."

Sans doute la probabilité dont nous par lons est relative à la règle adoptée, et quand on l'a appliquée après l'épreuve, on ne sait pas quelle chance on a d'avoir raison. Il est même assez difficle de donner un sens précis à cette question.

Mais ces difficultés ne sont nullement attachées à la question des intervalles de confiace.

Elles sont les memes quand, après une série de mesures, on donne la valeur numérique adoptée pour la grandeur et la precision qu'on attribue à ces mesures.

Une méthode de calcul d'une mesure n'est pas autre chose qu'une règle qui attache au résultat aléatoire d'une série de mesures un intervalle aléatoire, généralement déterminé par son centre et sa longueur.

La chance qu'on a d'avoir raison en disant que la grandeur est dans cet intervalle n'a pas de sens précis pour une épreuve unique, supposée faite.

Remarque. En réalité, ni la méthode de Bayes qui suppose connue la loi de probabilité *a priori*, ni la méthode des intervalles de confiance, qui ne suppose rien sur l'inconnue, ne conviennent généralement aux problèmes concrets.

On sait en général quelque chose sur l'inconnue que l'on cherche, mais ce quelque chose n'est pas aussi précis qu'une loi de probabilité *a priori*.

Et sans être très précis, c'est pourtant mieux que rien.

C'est une sorte de pondération intuitive du domaine ou peut se trouver l'inconnue.

Et il semble bien que cette pondération devienne une pondération *a posteriori* dont il n'est pas deraisonnable de supposer qu'elle se conduit à peu près conformément au théorème de Bayes.

Ces idées, auxquelles j'avais pensé depuis assez longtemps, ont été ravivées par des conversations avec Mr. Etienne Halphen que l'expérience statistique a amené à les approfondir je devrais dire à les découvrir et a leur donner une forme extrêmement nouvelle et originale.

En utilisant la formule de Bayes

Dans cet ordre d'idées p est une variable aléatoire qui a une certaine loi de probabilité *a priori*.

Cette loi *a priori*, assigne à tout intervalle compris entre $\alpha > 0$ et l'unité une certaine probabilité $P_0(\alpha)$.

Après la réussite d'une épreuve, la variable aléatoire p a une loi de probabilité nouvelle, liée par cette épreuve et qui est la loi de probabilité *a posteriori*.

L'intervalle considére précédemment a, dans cette nouvelle loi, une probabilité $P_1(\alpha)$.

On peut alors énoncer le resultat suivant: $P_1(\alpha) \geqslant P_0(\alpha)$, l'égalité n'étant pas exclue, mais n'étant vérifiée que si les deux probabilités sont nulles.

Autrement dit, la probabilité ne peut augmenter pour un intervalle qui comprend zéro, elle ne peut diminuer pour un intervalle qui comprend l'unité.

Pour plus de simplicité, supposons que $f(p)dp$ soit la probabilité élémentaire *a priori*.

La différence $$P_1 - P_0 = \frac{\int_\alpha^1 pf(p)dp}{\int_0^1 pf(p)dp} - \int_\alpha^1 f(p)dp$$

est nulle pour $\alpha = 0$ et pour $\alpha = 1$.

D'autre part sa dérivée a pour valeur

$$f(\alpha)\left[1 - \frac{\alpha}{A_0}\right] \qquad A_0 = \int_0^1 pf(p)dp = E_0(p)$$

Le résultat annoncé en résulte aisément.

La plus grande variation a lieu pour l'intervalle A_0, 1.

L'espérance mathématique *a posteriori* de p a d'ailleurs une proprieté analogue. Elle ne peut que croître. Mais cette fois, on a $A_1 > A_0$.

On peut dire encore que la courbe représantative de la fonction des probabilités totales ne peut que descendre après une épreuve favorable.

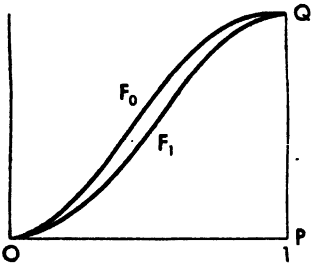


Fig. 1

La différence des aires, ou l'aire comprise entre F_0 et F_1 a pour valeur $A_1 - A_0 = \frac{\sigma_0^2}{A_0}$ en appelant σ_0 l'écart type *a priori* de p.

On voit que ce glissement de la courbe des probabilités totales vers le contour *OPQ* qui représenterait la certitude, correspond bien à l'idée intuitive que nous avons de l'effet d'une épreuve favorable.

Remarque. Si l'on veut étudier l'effet d'une épreuve complexe, résultant de deux épreuves successives, l'une favorable, l'autre défavorable, on peut signaler un résultat analogue. Cette fois, on peut prendre un intervalle ne contenant ni zéro, ni un. Si, pour plus de netteté, on prend un intervalle symétrique par rapport à $\frac{1}{2}$: $\alpha = \frac{1}{2} - h$, $\beta = \frac{1}{2} + h$ alors la probabilité, $P_1(\alpha\,\beta) \geqslant P_0(\alpha\beta)$. La concentration de probabilité, au lieu de se faire vers le point $p = 1$, se fait alors dans la région centrale.

La méthode precédente, parfaitement nette, exige pour être appliquée concrètement qu'on ait bien précisé la signification de la variable aléatoire p et de sa loi de probabilité *a priori*.

9. L. H. C. Tippett

SANS RECOURS A LA FORMULE DE BAYES

If an event E has an unknown probability p, and in *one* trial a *success* occurs, we infer with a certain degree of confidence that p is greater than a specific value p_0, the degree of confidence for each value of p_0 being measured by a "confidence coefficient" $\varepsilon = 1 - p_0$.

If an event E has an unknown probability p and in *one* trial a *failure* occurs, we infer that p is less than p_0 with a "confidence coefficient" of $\varepsilon = p_0$.

The interpretation of the confidence coefficient is as follows: if the above two rules of inference are applied to a large number, N, of trials, for each of which p may have a different value, the inference will be correct in a proportion u of the trials, where, in the statistical sense, u tends to ε as N tends to infinity.

10. M. Kendall

SANS RECOURS A LA FORMULE DE BAYES

La réponse que je donnerais serait basée sur la théorie des "*confidence intervals*" que j'ai traitée dans le second livre de mon "*Advanced Theory of Statistics*" (voir chapitre 19).

Le cas particulier $r = 1$, $n = 1$ est un peu spécial parce que la proportion observée ne peut atteindre que les deux valeurs 0 et 1. Mais les principes restent les mêmes. On choisit un niveau de signification (*significance level*), par exemple 5%. Les lignes de confiance deviennent 95% des axes des coordonées et des paralleles ainsi:

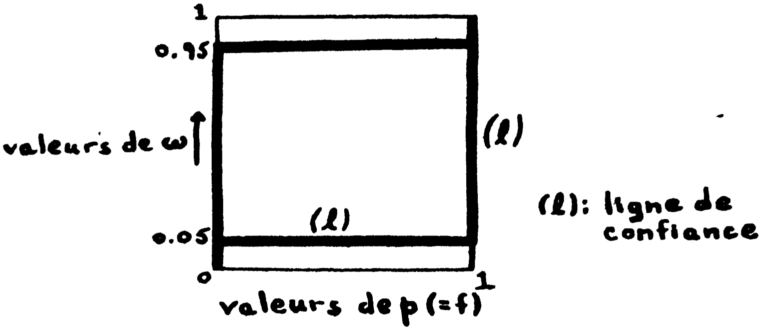


Fig. 2

Si l'on observe $f = 1$, l'ordonnée à $f = 1$, indique que l'on doit dire que $0.05 \leqslant \omega \leqslant 1$. Si $f = 0$ on dit que $0 \leqslant \omega \leqslant 0.95$.

Dans le sens de la théorie ces propositions seront réalisées dans 95 pour cent des cas.

D'un point de vue different, on pouvait dire qu'on devait rejeter toute hypothèse qui donnerait à l'événement une probabilité plus petite que 5%, ce qui conduit à la même conclusion; mais c'est un peu intutif.

Si l'on me forçait à attribuer un seul *nombre* à la valeur de ω, je m'appuierais sur le principe de *maximum likelihood*, et je dirais que $\omega = 1$, mais pour n si petit l'écart probable est si grand qu'il n'y a pas de justification pour une méthode d'estimation qui prétend fournir un résultat si précis,

11. R. C. Geary

SANS RECOURS A LA FORMULE DE BAYES

The Bureau's specific problem is solved as follows from the viewpoint of the well-known confidence interval theory. Nothing is known about the universal probability p of E so we must try to make a statement

about its magnitude which, if uniformly applied, will be incorrect in a proportion of cases to which an upper limit can be placed in advance. For the case $n = r = 1$ the procedure is as follows:

$$A \begin{cases} \text{If } E \text{ occurs, write } \omega \leqslant p \leqslant 1 \\ \text{If } E \text{ does not occur, write } \omega \leqslant q \leqslant 1 \end{cases} \begin{matrix} \text{with } p + q = 1 \\ \text{and } 0 < \omega < \tfrac{1}{2} \end{matrix}$$

It will be shown that if the experiment be repeated an indefinitely large number of times N the joint statement A will be incorrect in a proportion P of cases where $P < \omega$ whatever the frequency distribution of P itself in the universe.

Possible universes from which the sample could have been drawn can be divided into three categories; (1) those for which $p \geqslant \omega, q \geqslant \omega$, (2) those for which $p \geqslant \omega, q < \omega$, (3) those for which $p < \omega, q \geqslant \omega$. The case $p < \omega, q < \omega$ cannot arise because $p+q = 1$ and $\omega < \frac{1}{2}$. The following scheme shows the number of cases which can arise in each sub-category and the nature of the decisions.

	Number of cases	*Decision*
1) Universes $p_1 \geqslant \omega, q_1 \geqslant \omega$		
E	$\Sigma N_1 p_1$	Correct
not-E	$\Sigma N_1 q_1$	Correct
2) Universes $p_2 \geqslant \omega, q_2 < \omega$		
E	$\Sigma N_2\ p_2$	Correct
not-E	$\Sigma N_2\ q_2$	Incorrect
3) Universes $p_3 < \omega, q_3 \geqslant \omega$		
E	$\Sigma N_3\, p_3$	Incorrect
not-E	$\Sigma N_3\, q_3$	Correct
Total	$N = \Sigma N_1 + \Sigma N_2 + \Sigma N_3$	

In this scheme the Σ applies to all possible universes in each category. Nothing is known about the numbers N_i except that they are each indefinitely large and that their sum is N.

The proportion P of erroneous decisions is given by

$$p = (\Sigma N_2\, q_2 + \Sigma N_3\, p_3) \;/\; (\Sigma N_1 + \Sigma N_2 + \Sigma N_3)$$
$$< (\Sigma N_2 + \Sigma N_3) \;/\; (\Sigma N_1 + \Sigma N_2 + \Sigma N_3) \leqslant \omega$$

Hence we are assured that the proportion or probability of error P is less than $\omega \leqslant 1/2$. It is important, however, to note that this conclusion does not apply to E and to not-E cases *separately*. For example the proportion of incorrect decisions when E occurs will be P' given by

$P' = \Sigma N_3 p_3 \, / \, (\Sigma N_1 p_1 + \Sigma N_2 p_2 + \Sigma N_3 p_3)$ which is not necessarily less than ω for all values of the N_1, N_2, and N_3.

The joint statement A has been given by M. Fréchet.[10] The proof given above is, however, on somewhat different lines from Professor Fréchet's and possibly contains a different basic concept.

12. L. Féraud

SANS RECOURS A LA FORMULE DE BAYES

Pour commencer restreignons légèrement la *Question posée* et prenons un exemple concret.

Soit une urne contenant 10^8 boules parmi lesquelles B blanches et $(10^8 - B)$ noires. On a fait un tirage dans les conditions dites "au hasard." Il a donné une boule banche. Que peut-on dire de B?

On peut dire d'abord $1 \leqslant B \leqslant 10^8$ et à première vue il semble que l'on ne puisse rien dire de plus. Ce qui est vrai en un certain sens.

Toutefois, si l'on considère toutes les hypothèses admissibles

$$B = 1, 2, \ldots, 10^8$$

ne peut on accorder à certaines d'entre elles une préférence? Nous allons montrer dans quelles conditions et ce que l'on peut entendre par là.

Il reste à poursuivre la vérification de chacune des hypothèses admissibles. Supposons que l'une d'elles soit admise, $B = B_1$. Qu'en fera-t-on? On est alors devant un tirage au hasard dans une urne de composition connue. Décidons d'adopter le degré de probabilité $\alpha = 10^{-6}$, c'est-a-dire de considérer que tout ensemble auquel est affectée une probabilité inférieure ou égale à 10 sera *vide de resultats*.

Ainsi, au degré de probabilité 10^{-6} et pour un seul tirage: si $1 \leqslant B_1 \leqslant 10^2$ on conclura qu'il ne sortira pas une boule blanche, si $10^2 < B_1 < (10^8 - 10^2)$ on ne pourra tirer aucune conclusion, si $(10^8 - 10^2) \leqslant B_1$ on conclura qu'il ne sortira pas une boule noire.

Par suite, dans le premier cas et dans le premier cas seulement, la sortie d'une blanche, qui a été constatée, est en contradiction avec une conséquence de l'hypothèse. L'adoption de tout hypothèse définie par un $B_1 \leqslant 10^2$ est en contradiction avec le seul fait expérimental que l'on possède. Aucune des hypothèses $B_1 > 10^2$ ne conduit à une conclusion.

C'est pourquoi lorsqu'il s'agit de poursuivre la vérification des diverses hypothèses on accordera une préference à celles qui n'ont pas déjà rencontré un démenti, c'est-a-dire à la classe $B_1 > 10^2$. En d'autres termes, on a une raison pour commencer par poursuivre la vérification des hypothèses $B_1 > 10^2$.

---

[10] *Portugaliae Mathematica* (1946).

L'ensemble $10^2 < B \leqslant 10^6$ est *l'ensemble de confiance* (*confidence set*) dans la terminologie de Neyman et Pearson. Il va sans dire qu'il peut être beaucoup plus restreint et par conséquent beaucoup plus intéressant, beaucoup plus efficace que dans l'exemple schématique que nous venons de considérer.

Bien entendu, nous ne décidons pas entre les hypothèses admissibles d'une manière absolue, mais nous décidons, et encore provisoirement, seulement en vue de ce que nous voulons faire: se servir de l'hypothèse qui sera admise pour en tirer une conclusion ou degré de probabilité adopté 10^{-6}. Mais cela détermine l'action à entreprendre, au moins dans une certain mesure.

Le même raisonnement s'étend évidemment à la question telle qu'elle est posée pour $n = r = 1$ lorsque p n'est plus nécessairement une fraction de dénominateur connu. Si α est le degré de probabilité nous accorderons une préférence aux hypothèses définies par $\alpha < p < 1$. L'intervalle $(\alpha, 1)$ sera l'intervalle de confiance. Pour n, r quelconque, la partie du raisonnement qui nous intéresse ici reste la même: le problème a été traité par Clopper et Pearson[11] et les graphiques qu'ils ont établis permettent de déterminer aisément l'intervalle de confiance pour les degrés de probabilité 0.05 et 0.01.

13. O. Anderson

SANS RECOURS A LA FORMULE DE BAYES

It seems to me that your proposal "*d'examiner seulement le cas ou* $n=1$" goes too far. My answer to the more generalized question of $n > 1$ and $r \geqslant 1$ can be found on pp. 105-119 of my *Einführung in die mathematische Statistik* (Vienna, Julius Springer, 1935). It is preceded (but *not* suggested) by an article by Stanislas Millot, "*Sur la probabilité* a posteriori" (*Comptes rendus hebd. des Séances de l'Académie des Sciences*, Vol. 176, p. 30, Paris, 1923). Similar ideas are developed by V. Romanowsky and later on by L. Isserlis and R. Prigge.

Censorship regulations do not permit to send you a full or even shortened translation of the text in question. As to the law of large numbers and its consequence—the theory of confidence-intervals, I believe that they can be based only on the following non-mathematical lemma of A. Cournot: events with small mathematical probabilities occur seldom. The limit of confidence ω is different not only for different men but also for different degrees of importance of a failure. *E.g.*, the probability

---

[11] The use of confidence or fiducial limits illustrated in the case of the binomial (*Biometrika*, XXVI, 1934, p. 404).

1/100 of an imminent rain is regarded as nil, but the same probability of any serious typhus or cholera infection is by no means negligible.

14. J. B. D. Derksen

WITHOUT RECOURSE TO BAYES' THEOREM

The author agrees with the criticism raised by Mr. R. A. Fisher in his book *Statistical Methods for Research Workers* and in previous publications against the use made of the Bayes' Theorem for the solution of the so-called problem of inverse probability. Recurrence to Bayes' Theorem is often found in older textbooks on the theory of probability and also among certain present-day authors who seem to adhere to the old point of view.

It will be unnecessary to repeat the arguments raised by Mr. R. A. Fisher and other mathematical statisticians who have stressed the logical inconsistencies of the hypothesis underlying Bayes' Theorem. The proper solution to the problem how to derive conclusions about the unknown magnitude of a parameter specifying a universe if only a random sample of a finite number of observations are available, is given in the theory based on the notion of confidence intervals. This theory has been further developed by J. Neyman and E. S. Pearson and by others.

In the simple case stated in the inquiry, where $r = n = 1$, the confidence limits for the unknown value of p are $\varepsilon/2$ and 1. On the basis of the data of the random sample given, whereby $r = n = 1$, the only conclusion about the unknown magnitude of the parameter p is that it must satisfy the inequality $\varepsilon \leqslant p \leqslant 1$, if $1 - \varepsilon$ represents the so-called "fiducial probability." No probability statement can be attached to this inequality.

15. C. Jordan

Since contradictory opinions prevail on most subjects related to probability, it is necessary to state the definitions relative to what is meant by probability.

According to Ellis' empirical postulatum,[12] when we say that the probability of an event is equal to p, we mean that we expect in the long run, that the event will occur in a ratio nearly equal to p; from this it follows that we expect the event to happen $\nu = np$ times in n observations. Moreover if N sets of n observations are made, the ratio in each set is generally different and when we say that the *probability of the*

---

[12] Robert Leslie Ellis, "Foundations of the Theory of Probabilities" (*Transactions of the Cambridge Philosophical Society,* Vol, 8, 1842)

probability p is equal to ω, then we mean that the ratio $p = \nu/n$ will occur nearly $N\omega$ times in N future sets.

If we apply our probability-judgment to a single future event, then p is only an expression of the measure of the expectation of the future event. The greater it is, the more the phenomenon will be expected.

Our judgment concerning the probability of an event may be guided first by theoretical considerations; if there are several possibilities, which we consider as equally likely, some of them favorable to the occurrence of the phenomenon, and others adverse, then the greater the number of the first, the more probable the occurrence of the phenomenon will seem to us. Finally we shall say that the probability of the event is equal to the ratio of the number of favorable possibilities to the total number of possibilities, considered equally likely.

If the results of future observations conform to our expectation, then we conclude that our hypothesis concerning the equally likely possibilities was right.

Secondly, our judgment may derive from past observations. If an event has happened x_1 times in n_1 trials, then according to the empirical postulatum, we shall say that $p_1 = x_1/n_1$ is the most probable value of the probability p of the single event, that is we shall expect, that in the long run it will occur nearly in this ratio; for instance if in n observations it occurs x times, then x/n is nearly equal to x_1/n_1. This will generally happen if n and n_1 are large, if not then the ratios may be different.

The occurrence of a phenomenon whose probability is very small is not to be expected; but if it happens it is not necessarily surprising.

1. If the events $E_1, E_2, \ldots, E_s$ occurred respectively $x_1, x_2, \ldots, x_s$ times in n trials, the probabilities of the events being equal to $p_1, p_2, \ldots, p_s$, then the surprisingness[13] of this phenomenon may be measured by the quantity $\mathscr{S} = 1 - P_i/P_m$; where P_i is the probability of the system x_i observed and P_m that of the most probable system x_i. If $\mathscr{S}$ is near to zero the phenomenon is not surprising at all, if it is nearly equal to one, then it is very surprising, so that we must look out for an error or another explanation.

The *a posteriori* probability-judgment applied to a past single event concerns mainly its surprisingness.

2. Pearson's and Elderton's χ test. If starting from the hypothesis $p_1, p_2, \ldots, p_s$, the probability $\mathfrak{P}$ of a system x_i more probable than the observed one is small, then the hypothesis is to be accepted, if it

---

[13] Ch. Jordan, "Approximation and Test Criteria: χ test and Bayes' Theorem" (*Journal Société Hongroise de Statistique,* 1937).

is large, it must be rejected. Generally this will lead to the same result as the preceding method.

3. Fisher introduced the notion of the likelihood of the system p_i by

$$L = \Sigma x_i \log p_i$$

the greater the likelihood the greater the probability of the system p_i should be considered. Since antilog L is proportional to the probability of the system x_i given by *Bernoulli's* formula:

$$P(x_i) = \frac{n!}{x_1!\, x_2! \ldots x_s!}\, p_1^{x_1}\, p_2^{x_2} \ldots p_s^{x_s} \tag{1}$$

hence it is also proportional to the *a posteriori* probability of the system p_i, according to Bayes' theorem, if *a priori* every system p_i is considered equally likely. The likelihood is maximum at the same time as $P(x_i)$, that is if $p_i = x_i/n$ for every value of i. Denoting this maximum by L_m, the smaller the difference $L_m - L$ is, the likelier the system p_i.

4. The *a posteriori probability* of a system p_i corresponding to a hypothesis, given by Bayes, formula, divided by the *a posteriori* probability of the most probable system p_i is also a measure of the reliability of the hypothesis. Generally this method leads to the same result as the preceding.

Particular case: $n = 1, x_1 = 1, x_2 = x_3 = x_s = 0$. The probabilities $p_1, p_2, \ldots, p_s$ of the events are unknown. Conclusions concerning the hypothesis of a certain system $p_1, p_2, \ldots, p_s$ are required.

1. Surprisingness. If $p_i \geqslant 1/2$ then $P(x_i) = P_m$ and $\mathscr{S} = 0$. If $1/2 \geqslant p_i \geqslant 1/s$, we may still have $P(x_i) = P_m$ but not necessarily. The surprisingness of the occurrence of E depends not only on the probability p_i, but on the other probabilities too. For instance:

Hypothesis I. $p_i = 10/100$ for $i = 1, 2, \ldots, 10$. Since $\mathscr{S} = 0$ the occurrence of E is not surprising at all, and therefore the hypothesis may be accepted.

Hypothesis II. $p_1 = 10/100$, $p_2 = p_3 = \ldots = p_9 = 1/100$ and $p_{10} = 82/100$. Now $\mathscr{S} = 72 = 82$; the occurrence of E_1 is surprising enough, so that the hypothesis should be rejected.

2. *Pearson's* χ *theorem.* If $p_1 \geqslant 1/2$ then $\mathfrak{P} = 0$ and the hypothesis is to be accepted; if $p_1 \geqslant 1/s$ it is still possible that $\mathfrak{P} = 0$, as for instance in Hypothesis I; but it is not necessarily so. In the case of Hypothesis II, the probability of obtaining a more probable system than $x_1 = 1$ is equal to 0.82, therefore this hypothesis is to be rejected.

3. The *likelihood* of the system p_i in the particular case considered is $L = \log p_1$, it is independent of the other probabilities so that the likelihood of the two hypotheses above would be the same.

4. The *a posteriori* probability of the system p_i according to Bayes' theorem is also the same in both cases; therefore the first or the second method should be used.

Remark. If there are but two possibilities: the events E_1 and E_2, and if at the first trial the event E_1 happened, then the estimation probabilities of the system p_1, p_2 will give nearly the same result by the four methods.

If $p_1 \geqslant 1/2$ then $\mathscr{S} = 0$ and $\mathfrak{P} = 0$. If $p_1 < 1/2$ for instance $p_1 = 0.1$, $p_2 = 0.9$ then $\mathscr{S} = 1 - p_1/p_m = 8/9$ and $\mathfrak{P} = 9/10$. The likelihood of the system p_1, p_2 is equal to $L = \log 0.1$, the likelihood of the most probable system $L_m = \log 1 = 0$; antilog $L = 0.1$.

The *a posteriori* probability of p_1, p_2 by Bayes' theorem is $\mathscr{S}(p_1) = 2pdp$ and that of the most probable system $\mathscr{S}_m = 2dp$. Finally

$$\mathscr{S}(p_1) \,/\, \mathscr{S} = .1$$

Without the use of Bayes' Theorem

Epitome

The events $E_1, E_2, \ldots, E_s$ occurred respectively $x_1, x_2, \ldots, x_s$ times in n trials, the corresponding probabilities $p_1, p_2, \ldots, p_s$ are unknown. The acceptability of a certain hypothesis $p_1, p_2, \ldots, p_s$ is to be investigated. The probability of the system x_i observed, if the above hypothesis is accepted, is according to Bernoulli's formula

$$P(x_i) = \frac{n!}{x_1!\, x_2! \ldots x_s!}\, p_1^{\,x_1} p_2^{\,x_2} \ldots p_s^{\,x_s}$$

Denoting by P_m the probability of the most probable system x_i corresponding to the hypothesis, the surprisingness $\mathscr{S}$ of the occurrence of the system x observed will be defined by

$$\mathscr{S} = 1 - \frac{P(x_i)}{P_m}$$

If $\mathscr{S}$ is nearly equal to zero, the occurrence is not surprising at all, and the hypothesis of the probabilities p_i may be accepted. If $\mathscr{S}$ is near to one then the occurrence is very surprising, so that the hypothesis should be rejected.

In the *particular case* of $n = 1$ and $x_1 = 1, x_2 = x_3 = \dots = x_s = 0$ we have $P(x_1) = P_1$. If $p_1 \geqslant 1/2$ then $P(x_1) = P_m$ and $\mathcal{S} = 0$. If $1/2 > p_1 \geqslant 1/s$ then $P(x_1)$ may still be equal to P_m, but not necessarily; the surprisingness of the events depends on the other probabilities too. Example.

Hypothesis I. $p_1 = 10/100$ for $i = 1, 2, \dots, 10$. Hence $P(x_1) = p_1 = P_m$, hence $\mathcal{S} = 0$ and the occurrence of E_1 is not surprising at all.

Hypothesis II. $p_1 = 10/100, p_2 = p_3 = \dots = p_9 = 1/100$ and $p_{10} = 82/100$. Now $P(x_1) = p_1$ and $P_m = p_{10}$ hence $\mathcal{S} = 72/82$; the occurrence of E_2 is surprising enough so that the hypothesis should be rejected.

If $n = 2$ and $p_1 \geqslant 1/2$ then $\mathcal{S} = 0$. If $p_1 < 1/2$ then

$$\mathcal{S} = 1 - \frac{p_1}{1 - p_1}$$

16. H. Cramér

Allow me first to observe that the discussion on pp. 507-517 of my book *Mathematical Methods of Statistics* covers the essential parts of the question at issue, so that I may refer to this work.

Applied to the particular problem mentioned in the *Enquête*, the main points of principle involved in the above-mentioned discussion reduce to the following statements:

1) There are cases in statistical practice, where the application of Bayes' theorem is perfectly legitimate, and where there exists a well-determined *a priori* distribution of the unknown probability p. If, in one single observation, the event E (of unknown probability p) has been found to occur, Bayes' theorem then gives a determined value to the *a posteriori* probability that p lies in any given interval $p_1 < p < p_2$.

2) In many cases, perhaps even in most cases, occurring in statistical practice, the conditions for the application of Bayes' theorem are certainly not satisfied. In cases of this character, the calculation of an *a posteriori* probability according to Bayes' theorem is entirely unjustified.

3) Whether the conditions for a legitimate application of Bayes' theorem are satisfied or not, the theory of confidence intervals as developed by Neyman can always be applied. In the particular problem mentioned in the *Enquête*, this theory leads to the following results. Let ε be given such that $0 < \varepsilon < \frac{1}{2}$, and let us adopt the following rule:

When, in one single observation, the event E occurs, we shall assert that the unknown probability p satisfies the inequality $\varepsilon < p \leqslant 1$.

When, on the other hand, E fails to occur, we shall assert that $0 \leqslant p < 1 - \varepsilon$.

Then, whatever be the value of the unknown probability p, the probability of making a wrong statement will always be $\leqslant \varepsilon$.

4) The particular problem mentioned in the *Enquête* does not allow the conclusions of the kind specific to the theory of confidence intervals to appear in their simplest form, since it introduces the additional complication connected with a discontinuous distribution. In cases of this type, the probability of making a wrong statement cannot in all cases be made to coincide with the given number ε, but can only be rendered $\leqslant \varepsilon$ (*cf.* pp. 513-514 of my book).

Résumé

This is a report on an inquiry, established by the author, among various scientists in different countries, asking their opinion about the principles of statistical estimation.

The general problem of statistical estimation of parameters is only considered in a special case where mathematics will not hide the principles which are to be discussed: We know the frequency $f = r/n$ of a random event E in n trials. Admitting that E has a constant though unknown probability p in these n trials, what can we say about p? To make things even simpler, we will only consider the case where $r = n = 1$: Knowing that in one trial, E has occurred, what can we say about p? Of course, nothing, from the practical point of view, except that E is possible. But as principles only are questioned, and since those will be exactly the same when n is greater than one, can we say something more than $0 \leqslant p \leqslant 1$, even if the obtained precision has no value from the practical point of view? In this short summary, we will be able but to give opinions and judgements based on the answers to the inquiry, the full text of which is included in the report.

These answers are varied and conflicting. They examine the question from different points of view.

From the purely mathematical point of view, it appears that when we use exclusively the classical axioms of the Calculus of Probability, it is absolutely impossible to deduce from the fact that $f = 1$, any other information about p than $0 \leqslant p \leqslant 1$ (even $p = 0$ being possible). From the same purely mathematical point of view this statement might be pushed further and be stated in a case appearing even more paradoxical from the practical point of view: a point chosen at random on a segment might fall a million times successively on the middle of this segment though, each time, the probability of this event should be zero.

Thus, to derive information from the knowledge of f, about the value of p, it is necessary either to invoke a practical principle not belonging to the axiomatic Calculus of Probability or to modify (lightly in appearance) the considered problem without changing the hypotheses by modifying the question, or by modifying slightly the hypotheses.

We may suppose that there is a definite *a priori* probability law of a random probability P which is the considered trial was equal to p. In that case the Bayes-Laplace formula[1] enables us to answer the modified question: What is the *a posteriori* probability that $P = p$ (or $p < P < p+dp$) in the categories of trials where E was realized.

Generally the *a priori* probability law of P, when existing, is unknown, but often we know something in advance which enables us to have some information about p and from that we may derive, if not the *a posteriori* probability law of P, at least some information about this law.

Furthermore, several authors have proved that supposing some very general conditions to be fulfilled by the *a priori* probability law of P, the *a posteriori* law of P converges to one easy-to-define probability law of P when n increases indefinitely.

But many others do believe that in most practical circumstances not only is the *a priori* probability law of P unknown, but also that it has no meaning. There are two ways of escaping that difficulty.

First we may use a Cournot's lemma: an event which has a very small probability may practically be considered as impossible; or in a modified form suggested by a Divisia remark: in practical life if an event has a very small probability, we may act *as if* it were impossible. Starting from this and admitting that when an event has a probability smaller than $\omega > 0$ $\left(\text{for example, } \omega = \frac{1}{10^{20}}\right)$, we may act as if it were impossible, it is easy to prove that when E of probability p, has been realized in one trial, we may act as if $\omega \leqslant p \leqslant 1$, an inequality which solves the problem in principle (though from the practical point of view it does not differ essentially from $0 \leqslant p \leqslant 1$).

---

[1] While it is stated sometimes that Laplace restricted to the case of a uniform *a priori* probability law the general formula given by Bayes without supposing equal the *a priori* probabilities, it is just the reverse which is true (Bayes still preceding Laplace for the particular case).

In that inequality $1-\omega$ is the "confidence coefficient;" it is essential to keep in mind that nothing in the previous reasoning allows one to say that it is the probability that this inequality hold true. For, in this reasoning this probability is 0 or one, as p is no more what it was when applying the Bayes-Laplace formula, a random element.

Some authors strongly object to Cournot's lemma, although perhaps, it is true, they would accept its modified form. They may save the situation by adopting the following last point of view which is favored in most of the answers. In many cases estimation of parameters must be done by professional operators. And it is only wanted that these operators do not often make mistakes.

Then suppose each of them should apply the following R_ω rule:

When E occurs in one trial, it is estimated that $\omega \leqslant p \leqslant 1$
When E does not occur in one trial, it is estimated that
$0 \leqslant p \leqslant 1-\omega$

Then it can be proved that the probability that an operator makes a false estimation is smaller than ω whatever be the value of p, whether it be constant, chosen through a fixed rule, or random.

RAISONNEMENT INDUCTIF OU COMPORTEMENT INDUCTIF ? LES CONCEPTIONS MODERNES DE LA STATISTIQUE MATHEMATIQUE

par Jerzy Neyman
Directeur du Laboratoire de statistique
de l'Université de Californie (Etats-Unis)

1. On a exprimé l'opinion que le calcul des probabilités et la statistique mathématique forment la base d'un procédé mental qu'on nomme "raisonnement inductif." Si l'on convient à n'employer le mot "raisonnement" qu'à la description de procédés qui conduisent au savoir, le raisonnement ne saurait être que déductif. D'autre part, comme il a été signalé en 1937 [1],[1] il y a lieu d'employer le terme "comportement inductif." On entend par ce terme les règles, plus ou moins générales, d'ajustement de nos actions aux résultats d'observations limitées.

Le calcul des probabilités et la statistique mathématique jouent un rôle important dans la formulation de ces règles. Certainement il y a assez de raisonnement, mais, comme d'habitude, tout raisonnement est déductif.

2. Le progrès humain est du aux "permanences" ou plutôt à notre aptitude reconnaître les permanences dans les objets qui nous entourent, ainsi que dans les changements de ces objets que nous appelons les phénomènes.

Parmi les "permanences" qui ont été remarquées les premières, doivent être les dimensions des objets, au moins les dimensions de certains objets. Par conséquent les dimensions sont considérées comme des propriétés des objets et on a inventé des règles pour les mesurer. Lorsqu'on a reconnu que, dans certains cas, les dimensions changent, on a cherché à discerner les "permanences" dans ces changements. Pour un grand nombre de classes de phénomènes, les "permanences" discernées dans le passé ont été retrouvées ultérieurement. Ceci a créé l'habitude de régler nos actions en tenant compte de permanences qui, pour le moment, semblent être établies. C'est justement ce que j'appelle le comportement inductif. Souvent il conduit aux résultats désirés.

3. Au 17ième et au 18ième siècles une nouvelle classe de "permanences" a été découverte. On la doit aux joueurs de jeux de hasard sur le continent de l'Europe et aux francs-maçons en Angleterre. Les "permanences" discernées auparavant étaient les dimensions, les

---

[1] Les chiffres entre crochets carrés indiquent les références correspondantes enregistrées à la fin de cette communication.

distances, les poids, etc. Les permanences nouvelles s'attachaient à des expériences qui, considérées séparément, ne permettaient pas de prévoir les résultats. Néanmoins, on a remarqué que les fréquences relatives de résultats possibles de ces expériences, répétées un grand nombre de fois, restent à peu près invariables.

Le premier tricheur au jeu de dés, qui a pensé à piper ses dés pour faire amener plus souvent les chiffres désirés, a été l'auteur de la conception nouvelle que la fréquence relative avec laquelle un dé tombe sur une de ses faces est une "permanence"; c'est une propriété de ce dé qui est mesurable et constante tout comme ses dimensions ou son poids.

Les francs-maçons en Angleterre ont fait une découverte semblable concernant le décès de personnes du même âge qui vivaient dans les mêmes conditions.

Une fois ces permanences discernées, il a été relativement facile de construire les conceptions abstraites correspondantes, dont l'ensemble constitue la théorie des probabilités. Comme d'habitude, les conceptions nouvelles ont trouvé des applications utiles comme guides dans le comportement inductif. Telles semblent être les origines du calcul des probabilités et de la statistique mathématique.

4. Il me semble que la statistique mathématique est un chapitre du calcul des probabilités. Les problèmes statistiques sont ceux du calcul des probabilités qui se rattachent le plus étroitement aux problèmes du comportement inductif. Les exemples qui suivent servent à éclaircir la distinction un peu délicate entre les problèmes de probabilité au sens strict et ceux de la statistique mathématique.

La question que le Chevalier de Méré adressa à Fermat, sur la probabilité P pour qu'en un jeu de 24 coups de deux dés "bons" on amène le "double six" une fois au moins, est une question de probabilité au sens strict. Il est probable que la réponse de Fermat, à savoir: $P = 0{,}491$ a servi au Chevalier de Méré comme guide sur quoi parier, c'est-à-dire, comme une règle du comportement inductif. Dans tous ces cas où les dés étaient bons, il fallait parier contre le "double six".

Pour citer un exemple des problèmes de la statistique mathématique, nous allons envisager les doutes que le Chevalier pouvait avoir s'il avait à jouer avec des dés inconnus, dont l'honnêteté était suspecte. J'imagine qu'il pouvait observer un ou deux jeux de 24 coups sans jouer et se décider à jouer et à parier après avoir pris connaissance des résultats. Par exemple, il a pu adopter des règles de comportement inductif comme les suivantes:

(i) De ne parier *sur* le "double six" dans le deuxième jeu et les suivants que si le "double six" a été amené au moins une of is danlse premier jeu, et de parier *contre* le "double six" dans le cas contraire.

(ii) De ne parier *sur* le "double six" dans le deuxième jeu et les suivants que si le "six" a été amené neuf fois au moins pendant le premier jeu et de parier *contre* le "double six" dans le cas contraire. *N. B.*, Si les dés sont bons, alors 8 est l'espérance mathématique du nombre des "six" dans un jeu.

(iii) De parier *sur* le "double six" dans le deuxième jeu et les suivants si le "double six" a été amené au moins deux fois dans le premier jeu. De parier *contre* le "double six" dans le deuxième jeu et les suivants si le "double six" n'a pas apparu pendant le premier jeu. Enfin, si le "double six" a été amené une seule fois, de rester spectateur du deuxième jeu en lui appliquant la règle (1). [2] [3].

Il est probable que plusieurs règles semblables ont été inventées dans le passé. *Trouver les propriétés de telles règles de comportement inductif qui possèdent de telles propriétés données, est le problème de la statistique mathématique.*

Le comportement basé sur les règles (i), (ii), ou (iii) se distingue essentiellement du comportement dicté par la réponse de Fermat au Chevalier de Méré. Dans le cas présent, l'adoption d'une des règles (i)-(iii) déterminait les actions du joueur comme une fonction du résultat aléatoire du premier jeu ou des deux premiers jeux. Par conséquent, les actions elles-mêmes obtenaient un caractère aléatoire.

Généralisons: *la statistique mathématique a pour but d'étudier les propriéteé de telles règles de comportement inductif qui déterminent nos actions comme fonctions bien définies des valeurs observées de variables aléatoires.*

Il nous faut expliquer ce qu'on entend par les "propriétés des règles de comportement inductif." Si le Chevalier de Méré avait adopté une des règles ci-dessus, pouvait être pour une seule raison: pour augmenter la fréquence des cas où il aurait parié sur le cas favorable, c'est-à-dire, dont la probabilité est au moins un sur deux. Il s'en suit que les propriétés intéressantes des règles de comportement inductif sont celles relatives à la fréquence avec laquelle on peut atteindre les résultats désirés dans les cas différents pourvu qu'on adopte une telle règle.

5. En étudiant les problèmes multiples que je me permets d'attribuer au Chevalier de Méré, nous avons parlé de la plupart des conceptions de la statistique mathématique moderne. Pour préciser, désignons par une seule lettre $\mathfrak{X}_0$ la totalité des variables aléatoires observables. Les valeurs de ces variables vont être interprétées comme les coordonnées d'un point aléatoire X_0 dans un espace W. Soit x une des positions possibles de X_0 dans W. Soit C une classe de sous ensembles w de W sur laquelle la probabilité $P\{X_0 \epsilon w\} = \Phi_0(w)$ se trouve définie. Nous allons considérer le cas où la fonction $\Phi_0(w)$ est inconnue sauf qu'elle appartient à une famille F de fonctions de répartition des

probabilités $\Phi(w)$, toutes définies sur la même classe C des sous ensembles w.

Décrivons la situation d'une manière différente. Outre la variable X_0 dont les valeurs particulières seront fournies par l'observation, on considère toute une famille, soit F_x, de variables aléatoires X_Φ. Toute variable X_Φ définit la position du point aléatoire dans le même espace W et la probabilité pour que ce point-ci tombe dans un sous ensemble $w \in C$ est donnée par la fonction $\Phi(w) \in F$. On sait le variable X_0 appartient à la famille F_x, on ne connait pas son identité. C'est une autre caractéristique des situations considérées que, comme c'était le cas de de Méré, on est forcé de faire un choix entre des actions appartenant à un ensemble A d'actions a possibles. Si l'identité de la variable X_0 nous était connue, les circonstances du problème auraient dicté le choix d'action d'une manière bien claire. De plus, supposons que pour toute fonction $\Phi(w)$ de la famille F et pour toute action possible a de l'ensemble A une fonction $\Psi(a/\Phi)$ nous est fournie qui représente le "gain" qui aurait résulté de l'adoption de cette action a dans les cas où la fonction $\Phi(w)$ représente la loi de probabilité de la variable X_0.

Soient $\Phi_1(w) \in F$ et a_1 et a_2 deux actions de l'ensemble A telles que $\Psi(a_1|\Phi_1) < \Psi(a_2|\Phi_1)$. Alors, si $\Phi_1(w)$ representé la loi de probabilité de X_0, l'action a_2 est préférable à l'action a_1. La fonction $\Psi(a|\Phi)$ peut être appelée la fonction de préférence des actions.

Dans le cas du Chevalier de Méré la variable aléatoire observable X_0 peut être identifiée avec le système de 24 variables T_i ($i = 1, 2, \ldots, 24$). Pour le i-ième coup de deux dés, définissons la variable T_i comme le nombre des "six" amenés par ce coup. C'est-à-dire, les valeurs possibles de T_i sont O, 1 et 2. L' espace W est composé de 3^{24} "points," chacun à 24 coordonnées, soit $t_1, t_2, \ldots, t_{24}$ toutes capables des valeurs zéro, un et deux. Chaque point x de W représente un résultat possible d'un jeu de 24 coups; donc il représente une position possible du point aléatoire X_0.

Soient $p_1{}^\circ$ et $p^\circ{}_2$ les probabilités d'amener un "six" avec le premier et le second dé respectivement. En adoptant l'hypothèse plausible d'indépendance, la probabilité pour que X_0 possède une position déterminée x être écrite:

$$(1)\quad p\{X_0 = x\}$$

$$= \left[(1-p_1{}^\circ)(1-p_2{}^\circ)\right]^{n_0}\left[p_1{}^\circ(1-p_2{}^\circ)+p_2{}^\circ(1-p_1{}^\circ)\right]^{n_1}\left[p_1{}^\circ\, p_2{}^\circ\right]^{n_2}$$

où $n_0+n_1+n_2 = 24$. Cette probabilité définit d'une manière évidente la fonction de répartition des probabilités $\Phi_0(w)$ de la variable X_0. Si l'on ignore les valeurs de p°_1 et p°_2 excepté que ce sont des nombres positifs plus petits que un, alors on est conduit à considérer toute une famille F de lois de répartition $\Phi(w)$ qu'on peut maintenant désigner par $\Phi(w \mid p_1 p_2)$ toutes étant définies par la formule

$$\Big[(1-p_1)(1-p_2)\Big]^{n_0}\Big[p_1(1-p_2)+p_2(1-p_1)\Big]^{n_1}\Big[p_1\,p_2\Big]^{n_2}$$

dans laquelle p_1 et p_2 désignent des nombres arbitraires entre zéro et un. Par conséquent, la famille F est, dans le cas considéré, une famille de fonctions à deux paramètres p_1 et p_2. Si l'on veut on peut aussi, à la place de la famille F, considérer une famille correspondante F_X de variables $X_{p_1 p_2}$ dont les lois de probabilité sont données par $\Phi(w \mid p_1 p_2)$. Donc, par exemple, $X_{\frac{1}{5},\frac{1}{7}}$ désigne le résultat aléatoire d'un seul jeu à 24 coups de deux dés dont le premier a une probabilité 1/5 d'amener un " six " et le second une probabilité 1/7. On ne sait rien sur la variable X_0, excepté qu'elle appartient à la famille F_X. Le Chevalier de Méré se trouvait confronté avec la nécessité du choix entre deux actions possibles: a_1 de parier *contre* le "double six" ou a_2 de parier *sur* le "double six". Donc, dans ce cas, l'ensemble A d'actions possibles ne possède que deux éléments a_1 et a_2. Pour tout élément $\Phi(w \mid p_1 p_2)$ de la famille F la fonction de préférence $\Psi(a \mid \Phi)$ doit être définie pour ces deux actions seulement. Si l'on considére les détails des règles du jeu, la définition de $\Psi(a \mid \Phi)$ peut être assez compliquée. Pour simplifier, définissons $\Psi(a_1 \mid \Phi_{p_1 p_2}) = 1$ et $\Psi(a_2 / \Phi_{p_1 p_2}) = 0$ si $1-(1-p_1 p_2)^{24} \leqq \frac{1}{2}$ avec les valeurs évidentes dans le cas contraire.

Retournons au cas général et considérons la règle R du comportement inductif comme suit: *si l'observation donne $X_0 = x$ l'action adoptée sera une action* $a(x)$, où $a(x)$ est une fonction définie pour tout $x \in W$ dont les valeurs sont dans A. La fonction $a(x)$ s'appelle la fonction de décision statistique. Les règles (i) et (ii) considérées ci-dessus sont des exemples de la règle R. Pour la règle (i) la fonction $a(x) = a_1$ toutes les fois qu'aucune des coordonnées de x est égale à 2.

Pour toute fonction $\Phi(w)$ de la famille F définissons une variablealéatoire, soit

$$Y_\Phi = \Psi[a(X_\Phi) \mid \Phi].$$

L'ensemble des variables Y_Φ ou plutôt l'ensemble de leurs lois de répartition de probabilité est appelé *la caractéristique opérationelle de la règle R* du comportement inductif.

Dans le cas du Chevalier de Méré, la caractéristique opérationelle de la règle (i) est une fonction de p_1 et p_2 et représente la probabilité qu'en pariant selon la règle (i), on parie sur le cas dont la probabilité est *plus forte que $\frac{1}{2}$. Il est évident què connaître les propriétés d'une règle de comportement R est connaître sa caractéristique opérationelle.*

6. *Le terme "règle de comportement inductif" suggère bien la conception pour laquelle il a été inventé, mais il semble aussi prétenctieux que le terme "raisonnement inductif." Donc il pourrait être préférable* d'employer le terme "procédé statistique" pour désigner l'ajustement de nos actions aux valeurs observées des variables aléatoires.

7. Les premières règles du comportement inductif ou les premiers procédés statistiques dont la théorie a été étudiée sont ceux dont la fonction de décision $a(x)$ n'a que deux valeurs. Les procédés statistiques de ce genre sont connus comme les procédés de vérification des hypothèses statistiques. La caractéristique opérationelle d'un procédé de vérification est connue comme sa fonction de puissance. [4] [5] [6].

Presque simultanément la théorie d'un autre genre de procédés statistiques a été étudiée, celui de l'estimation statistique au moyen d'un intervalle. Cette théorie est connue comme la théorie des intervalles de confiance. [7] [1].

Ces deux théories, celle de vérification des hypothèses statistiques et celle des intervalles de confiance ne sont que des cas particuliers de la théorie générale des procédés statistiques (= des règles du comportement inductif). Cette théorie générale est due à Abraham Wald. [8] [9]. Dernièrement un grand nombre de travaux importants ont été publiés se rattachant aux divers aspects de ces théories, mais je ne puis citer que quelques auteurs.

Littérature

[1] J. Neyman, "L'estimation statistique traitée comme un problème classique de probabilité," *Actualités Scientifiques et Industrielles*, N. 739 (1938), pp. 25-57.

[2] Walter Bartky, "Multiple Sampling with Constant Probability," *Annals of Mathematical Statistics*, Vol. 14 (1932), pp. 363-377.

[3] Abraham Wald, "Sequential Tests of Statistical Hypotheses," *Annals, of Mathematical Statistics*, Vol. 16 (1945), pp. 117-186.

[4] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society of London*, Vol. 231-A (1933), pp. 289-337.

[5] J. Neyman and E. S. Pearson, "Contributions to the Theory of Testing Statistical Hypotheses," *Statistical Research Memoirs* Vol. 1 (1936), pp. 1-37 ; Vol. II (1938), pp. 25-57.

[6] J. Neyman, "Sur la vérification des hypothèses statistiques composées," *Bulletin de la Société de Mathématique de France*, Vol. 63 (1935), pp. 246-266.

[7] J. Neyman, "Outline of a theory of Statistical Estimation Based on the Classical Theory of Probability," *Philosophical Transactions of the Royal Society of London*, Vol. 236-A No. 767 (1937), pp. 333-380.

[8] Abraham Wald, "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *Annals of Mathematical Statistics*, Vol. 10 (1939), pp. 299-326.

[9] Abraham Wald, *Sequential Analysis*, J. Wiley and Sons, New York, 1947.

[10] H. F. Dodge and G. Romig, "A Method of Sampling Inspection," *The Bell System Technical Journal*, Vol. 8 (1929), pp. 613-631.

[11] M. A. Girshick, "Contributions to the Theory of Sequential Analysis," *Annals of Mathematical Statistics*. Vol. 17 (1946), pp. 123-143, 282-298.

[12] Harold Hotelling, "Some Improvements in Weighing and Other Experimental Techniques," *Annals of Mathematical Statistics*, Vol. 15 (1944), pp. 297-306.

[13] P. L. Hsu, "On the Power Function of the E^2-Test and the T^2-Test," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp. 278-286.

[14] A. Kolmogoroff, "Confidence Limits for an Unknown Distribution Function," *Annals of Mathematical Statistics*, Vol. 12 (1941), pp. 461-463.

[15] R. V. Mises, "On the Classification of Observation Data into Distinct Groups," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp. 68-73.

[16] H. B. Mann, "On a Test of Randomness Based on Signs of Differences," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp. 193-199.

[17] E. S. Pearson, "The Probability Integral Transformation for Testing Goodness of Fit and Combining Independent Tests of Significance," *Biometrika*, Vol. 30 (1938), pp. 134-148.

[18] E. S. Pearson, "The Choice of Statistical Tests Illustrated on the Interpretation of Data Classed in a 2 x 2 Table," *Biometrika*, Vol. 34 (1947), pp. 139-167.

[19] E. J. G. Pitman, "The Closest Estimates of Statistical Parameters," *Cambridge Philosophical Society Proceedings*, Vol. 32 (1937), pp. 212-222.

[20] Henry Scheffé, "On Solutions of the Behrens-Fisher Problem Based on the *t*-Distribution," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 35-44.

[21] Charles Stein, "A Two-Sample Test of a Linear Hypothesis Whose Power is Independent of the Variance," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp, 243-258.

[22] B. L. Welch, "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved," *Biometrika*, Vol. 34 (1947), pp. 28-35.

[23] S. S. Wilks, "Statistical Prediction With Special Reference to the Problem of Tolerance Limits," *Annals of Mathematical Statistics*, Vol. 13 (1942), pp. 400-409.

[24] J. Wolfowitz, "Additive Partition Functions and a Class of Statistical Hypotheses," *Annals of Mathematical Statistics*, Vol. 13 (1942), pp. 247-279.

Résumé

In the past, claims have been made that mathematical statistics and the theory of probability form the basis of some mental process described as "inductive reasoning." If it is agreed to limit the meaning of the word "reasoning" to those mental processes which lead to knowledge, then all reasoning seems to be deductive. On the other hand, as was first remarked in 1937, there seems to be room for the term "inductive behavior." This may be used to denote the tentatively general rules which adjust our behavior to limited amount of observation.

In establishing these rules, both the theory of probability and statistics do play an important role, and there is a considerable amount of reasoning involved. However, as usual, the reasoning is all deductive.

Human progress is based on permanencies or, rather, on our ability to detect permanencies both in the objects surrounding us and in changes in those objects which we describe as phenomena.

With many phenomena the permanencies observed in the past often were noted in the same classes of phenomena observed later. This created the habit of regulating our actions by referring to the established permanencies. This is inductive behavior.

In the 17th and the 18th centuries, a new kind of "permanency" was detected by gamblers on the continent of Europe and by Masons in England. Previously detected permanencies were dimensions, distances, weights, etc. The newly detected permanency was the relative frequency with which a particular result recurs in repeated trials in which the outcome of any single trial is unpredictable.

The first gambler who thought of "loading" a die is the author of the concept that the relative frequency with which the die falls with six dots up is this die's permanency.

Once these permanencies were noticed, the corresponding abstract concepts were easy to create. The postulated long run relative frequency received the label of probability and found useful applications in forming our inductive behavior. This was the origin of the calculus of probabilities and of mathematical statistics.

The theory of mathematical statistics is part of the theory of probability and statistical problems are most directly generated by the problems of inductive behavior.

Discussion

Mr. Gini suggested to Mr. Neyman that mathematical statistics should not be restricted to being a segment of the calculus of probability. He noted further that the "permanences" exist independently, and that he thought it unnecessary or inappropriate to consider the applications of probability theory to statistics as rules which deal with inductive behavior.

Mr. Gini then turned to the question of accounting for past experience in ascribing preference to one or another hypothesis. He stated that the means of verifying the hypothesis suggested by Mr. Neyman takes account only of present experience, and thought, therefore, that such a method should be applied only when there is no past experience. He refused to proceed as if the *a priori* probability did not exist, and saw a danger in the application of the modern theories of probability simply to get conclusions at any cost.

Mr. Hotelling pointed out the influence of philosophical pragmatism on the discussion of problems of inference. By this pragmatic subordination of the idea of belief to the idea of action, statisticians have turned to questions of determining which of alternative actions to take (acceptance or rejection of a hypothesis, selection of statistical methods of making decisions). Mr. Hotelling believes that only through the emphasis on action can statisticians make great progress where there is a multiplicity of available decision or action. Mr. Hotelling noted that cases of multiple decisions have naturally and almost necessarily led to repeated experiments and samplings, to sequential analysis.

Mr. Hurwicz commented that the divergency of views of Messrs. Neyman and Gini is only apparent, pointing out that Mr. Neyman's procedure provides in a particular problem for a class of admissible distribution functions, whose choice will be dictated by previous knowledge.

Mr. Rubin discussed the use of previous data, stating that they cannot be rigorously treated unless the mechanisms by which earlier conclusions were arrived at are known. An *a priori* probability cannot be set up because no probability for the original hypothesis can be derived from the original data.

Mr. Bernstein pointed out that exclusion of *a priori* probabilities would render impossible much work in medical statistics and genetics, where the experience of laboratory genetics has given indications about the *a priori* probability of the hypothesis when data concerning humans are limited.

Mr. Bernstein continued, to oppose the viewpoint of Messrs. Neyman and Hotelling that this purpose is to find rules of behavior. He statde his belief that the human observer is a secondary element in the study of nature, and that there is no justification to be drawn from pragmatism for a changed attitude throughout natural science.

Mr. Eisenhart proposed two examples to focus attention on the fact that more than the favored hypothesis must be considered, that the function of a statistician is to give advice in the use of data as a basis for decision. In both, one is led to the formulation of statistical tests in terms of the admissible decisions and acceptable risks of errors.

Mr. Nanda illustrated, with a discussion of sampling for agricultural information, the ways in which past information may be used in the design of samples.

Mr. Marschak added a few remarks on the development of theoretical economics and the preoccupation with the theory of choice (*i.e.*, rational behavior or what happens if people try to maximize their utility). This utility is described as a function of certain actions.

Mr. Tukey said that the statistician cannot often enter into the question of deciding whether a given theory fits the facts well enough for present purposes—a viewpoint adopted in physical or biological science where it is expected that hypotheses will be modified. The problem of statistics is to guess what methods will work out well where the hypothesis used to set up the methods is not exactly true.

Mr. Hurwicz foresaw a difficulty in enforcing one element of the procedure described by Mr. Neyman as statistical inference, namely that the action must be prescribed by the rule, which is a well-defined function of observable random events.

Mr. Wald replied briefly to a question put by Mr. Marschak on the decision problem and the minimax principle as a rule for behavior.

Mr. Fréchet proposed a hypothesis which might reconcile those who will not take into account the application of the Bayes-Laplace formula. He also expressed the opinion that there is a place both for the scholar who wishes to understand what has happened, and for the statistician whose function is to determine courses of action.

Mr. Neyman commented that his views should not be interpreted to mean that *a priori* probabilities should not be utilized in such cases (as genetics) where they have a meaning, and stated that the question of principle under discussion is whether the theory of rules of inductive behavior has meaning. A further question is whether cases exist where there is no *a priori* probability, or where the *a priori* probability is unknown. Mr. Gini reaffirmed his views that such cases are very infrequent, and added that even in the cases of multiple decision the *a priori* probability of each hypothesis must be considered. He concluded with the prediction that after further study the divergent views will be reconciled.

STATISTICAL METHODOLOGY SECTION

Monday, September 15 at 2:00 p.m.

Experimental Design in the United Kingdom and the United States

CHAIRMAN :

R. A. Fisher

DEVELOPMENT OF THE THEORY OF EXPERIMENTAL DESIGN

by R. A. Fisher
University of Cambridge (Great Britain)

At the time, now about thirty-five years ago, when first I came to study statistical methods, nothing was further from my thoughts, or from those of my contemporaries, than that the art of experimental design would ever come to be, as it now surely is, an integral part of the subject. The horizon was indeed filled with problems of *fitting*, or as we should now say, of estimation; and the two primary tests of significance, Pearson's χ^2 test, and Student's test, though certainly not widely understood, were already available, though in the case of χ^2 still in a form that ignored the degrees of freedom. The mathematical theory of all these aspects of statistical methodology was frankly in chaos. No step forward was possible without concurrently clearing up a number of errors and misapprehensions, only indirectly connected with the matter in hand.

The line of statistical thought which led most directly to the problems of experimental design, was undoubtedly the Theory of Estimation, devoted to the study of the distributional properties, in appropriately chosen sets of samples, of the various possible estimates, or estimators, which may be used to estimate the unknown parameters in the problem. Even though at this stage of discussion no thought was given to the manner in which the data were obtained, but these data were considered as an immutable part of the problem, yet it gradually became clear that absolute criteria existed by which one estimate might be judged, without ambiguity, to be better than another, and that in the Method of Maximum Likelihood the practical worker possessed a tool providing estimates which even in theory could not be improved upon.

In the course of this work it was shown that one valid method for the comparison of estimates was supplied by the amount of information relevant to the unknown parameters, this quantity being a scalar for a single parameter, and a symmetrical matrix in general. It was rigorously demonstrated that the amount of information extracted in the process of estimation could never exceed the quantity supplied by the data.

This finding, combined with the practical fact that directly available processes of computation would extract almost always a very large fraction of the total available, shifted the *moral* balance. The statistician was no longer responsible for the accuracy or precision of the results of his labors. His business became much less like that of a conjurer who is expected to work wonders, and more like that of a chemist who undertakes to assay the proportion of gold in a sample of ore. He need not be ashamed if the assay is low, or elated with pride if it is high. Consequently, provided he used available methods of proved competence, and did his arithmetic right, the statistician had no longer any responsibility for the value, or for the worthlessness, of his estimates. The weight of his responsibility was thrown back on to the process by which the data had come into existence. So, armed with the concept of amount of information as a practical tool, statisticians came to study what forms of experiment, or what types of observational programs would yield the most information for a given expenditure in time, money, and labor.

The purpose of these introductory remarks of mine will have been served if I have made clear the logical necessity of this step in the history of the subject, by which interest shifted from the problem of minimizing loss of information in the processes of statistical reduction, to the problem of maximizing the amount of information created by a well planned expenditure of experimental resources. There is but one further point I need make.

The only modification of the art of experimentation, as developed on statistical principles, which does not flow directly from the requirement to maximize the amount of information, obtainable with limited resources, consists in the introduction into the experimental procedure of what is called *randomization*. The theory of estimation presupposes a process of random sampling. All our conclusions within that theory rest on this basis; without it our tests of significance would be worthless. In order to justify the conclusions of the theory of estimation, and the tests of significance as applied to counts or measurements arising in the real world, it is logically necessary that they too must be the results of a random process. In controlled experimentation it has been found not difficult to introduce explicit and objective randomization in such a way that the tests of significance are demonstrably correct. In other cases we must still act in the faith that Nature has done the randomization for us, as for example, when we conduct a series of variety trials over a period of five years, and draw our conclusions *as if* the weather experienced in these years were really a random sample of five out of an infinite population of seasons determined by the climate prevailing at the site of experimentation. In all cases, however, we now recognize randomization as a postulate necessary to the validity of our conclusions, and the modern experimenter is careful to make sure that this postulate is justified.

My few remarks, Ladies and Gentlemen, are aimed at displaying the seed and roots of this young plant; later speakers will I hope tell us something of its flowers and fruit.

Résumé

L'auteur expose le développement des idées ayant conduit, via la Théorie de l'Estimation, à l'étude des plans d'expériences.

On s'approfondit d'abord sur le problème de minimiser la perte d'information dans le procès de réduction d'un ensemble d'observations, plus tard sur le problème de rendre maximum la valeur informative résultant d'une utilisation aussi efficient que possible des ressources disponibles. L'on a introduit aussi dans la procédure expérimentale l'exigence de fortuité.

Discussion

Mr. Bose presented the following brief commentary on the work in experimental design which has been carried on in India, in the Indian Statistical Institute and the Calcutta University.

"...The three basic principles of 'experimental design' are randomization, 'replication' and 'local control.' As Fisher has remarked, design and analysis are two aspects of the same logical whole, and randomization and replication are necessary in order that valid tests of significance, based on the use of t and z tests should be possible. In order, however, to increase the efficiency of the experiment, local control is necessary. The main principle is to divide the experimental material into relatively homogeneous parts called 'blocks.' In the well known randomized block design every treatment is applied to one randomly chosen experimental unit within each block, and the analysis is so arranged that the block effects are eliminated in making treatment comparisons. When, however, the number of treatments is large, then the differential responses of the experimental units within the same block swell the error, and make the experiment inefficient. We have thus to find a way of reducing the block size, and this has given rise to the device of 'confounding', the use of 'incomplete block design.' The incomplete block designs most commonly in use are the 'lattice designs' and the 'balanced incomplete block designs.' In the latter, v treatments have to be arranged in b blocks of k experimental units each so that treatment is replicated r times, and each pair of treatments occurs together in λ blocks (the number λ being independent of the pair of treatments we start with). It is readily seen that

$$bk = vr, \qquad \lambda(v - 1) = r(k - 1)$$

and Fisher has shown that

$$b \geqq v$$

"Given the parameters v, b, r, k, λ, satisfying the above conditions, the combinatorial problem of actually producing the required arrangement still remains. The practically useful cases are when $r \leqq 10$ and a number of these were listed in Fisher and Yates tables. But many blacks remained. Indian workers have succeeded in filling up most of these blanks; some of the new solutions were included in the second edition of Fisher and Yates table, and other are going to be included in the forthcoming third edition.

"The total number of available balanced imcomplete block designs is unfortunately very restricted. To a certain extent the deficiency is made good by lattice desings, in which though every pair of treatment is not compared with the same accuracy, the analysis is simple and straightforward. In order to fit with various experiments which arise in practice, a much more flexible system of designs is required. A more general system of designs, *viz.*, the partially balanced design, has been worked out by the statisticians of the Calcutta School, in which the stringent combinatorial restrictions to which balanced incomplete block designs are subject have been slightly relaxed, without losing the advantage of simplicity of analysis. The number of the these designs (not all of which have yet been worked out) is very large, and gives ample scope to the experimenter to choose a design according to his requirements.

"A special feature of the theoretical work done in India on the construction of designs has been the consistent use of methods and concepts of Modern Algebra, especially the theory of Galois Fields and Finite Geometrics. For example, a finite geometrical representation, has been found extremely useful in connection with the theory of the symmetrical factorial design, where its use easily leads to an exhaustive enumeration of all the possibilities of confounding in a given situation.

"Much work on 'asymmetrical factorial experiments,' 'two-way arrangements,' both factorial and non-factorial, and separation of complete replications (in the case of incomplete block designs), had also been done in India, but it is impossible in this short time to give details..."

Mr. Rasch gave his attention to two points which he felt deserve more consideration than they usually receive, and which show how biologists and statisticians may gain by yielding to each other's views:

(1) Choice of experimental object. The decision on the fundamental question: What is to be observed? is too often left to the common sense of the biologists. Dr. Rasch cited an experiment in which the collaboration between biologist and statistician was highly profitable and which pointed to the desirability of using preliminary studies with the aim of securing effective observations. He continued:

" ..But would it be possible to develop a Theory of Design of Preliminary Experiments? Undoubtedly! The alternative possibilities must, of course, be selected by the biologist after a discussion of the importance of the biological specifity of the reactions recorded. But a theory for the choice between the alternatives offered can be elaborated on the basis of the theory of reaction curves.

"To give a brief indication of an approach to the problem I shall assume the reaction curve to be linear with slope β and the variation of the observations about the reaction line to be normal with the variance σ^2. The mean error of the usual estimate of the mean reaction dose is $\frac{\sigma}{\beta}$ times a factor largely depending on the allotment of the animals to the doses.

Therefore, if the computation of that estimate is the main purpose of the experiment, the observation with the least value of σ should be preferred."

(2) Interpretation of systematically arranged experiments. Mr. Rasch described a simple serial arrangement which has been used in Denmark. He believed that when there must be some definite cause of variation, randomization is not to be preferred to research intended to reorganize that cause so that it may be accounted for in a systematic design.

Mr. Fisher commented on Mr. Rasch's remarks on systematic design in agriculture, that his process of fitting a polynomial in order to remove residuals might also be done with a randomized arrangement. Thus, this cannot be described as an advantage of a systematic over a random arrangement.

APPLICATION OF THE THEORY OF EXPERIMENTAL DESIGN IN BIOLOGY

The Utilization of Experimental Data for Improving Design

by George W. Snedecor

Statistical Laboratory, Iowa State College (United States)

The results of a well designed and well conducted experiment contain information not only on the subject being investigated but also on the improvement of the design for further trials. After extracting from an experiment the evidence which is immediately desired, the investigator is prone to abandon the data, proceeding to further experimentation without utilizing the knowledge that could be had about improvement of the design itself. Some even go to the extreme of repeating the same experiment time after time, only to learn at the end that the plan was inefficient or inadequate. This is wasteful of both time and money. The purpose of this paper is to illustrate the method of examining two experiments in which the design turned out to be susceptible to improvement.

The first illustration is based on a factorial experiment to learn the ascorbic acid content of snapbeans stored for four periods of time (2, 4, 6, and 8 weeks) at each of four temperatures (0°, 5°, 10°, and 20° F.). The beans to be used were all harvested under uniform conditions before eight o'clock one morning. They were prepared and quick-frozen before noon of the same day. Three packages were assigned at random to each treatment. The twelve packages to be kept at 0° F. were placed at random in the storage compartment, three being removed for analysis at the end of each storage period. In addition, twelve packages were quick-frozen, then tested at once for their ascorbic acid content. On each package there were two determinations of the concentration of the vitamin: The averages are recorded in Table 1.

TABLE 1

Average Concentration of Ascorbic Acid in Snapbeans (mg. per 100 g.)

Storage Time (Weeks)	Storage Temperature (Fahrenheit)			
	0°	5°	10°	20°
0	16.2			
2	15.0	14.2	14.9	11.4
4	15.7	16.0	14.3	9.2
6	15.4	14.4	13.7	7.0
8	15.5	13.2	12.3	5.5

The two sets of regressions, on storage time and on temperature, were examined for curvilinearity. The former were linear, starting at a common level at the initiation of the experiment and tending downward with their negative slopes increasing with temperature. The latter were parabolic, starting at a constant vitamin concentration at 0° F. (unaffected by time of storage) but with rate of loss increasing with duration of storage. These considerations led to the adoption of the following surface as the empirical model

$$Y = \alpha + \beta ST + \gamma ST^2 + e, \tag{1}$$

where Y, S, and T represent respectively ascorbic acid concentration, length of storage period, and temperature of storage.

Since three packages were tested at each temperature-storage treatment, the sampling variance was available at 17 points on the surface. This variance turned out to be homogeneous, unaffected by either temperature or time. This warranted the use of equal weights in making the least squares estimates of the regression constants. The fitted equation was

$$Y = 15.55 - 0.01437ST - 0.002735ST^2, \tag{2}$$

with $R^2 = 0.8995$. The analysis of variance is summarized in Table 2.

TABLE 2

Analysis of Variance of Ascorbic Acid Concentration in Snapbeans

Source of Variation	Degrees of Freedom	Mean Square
Total	119	
Regression Constants	2	531.2
Deviation of Means from Regression	14	4.14
Deviation of Packages from Means	43	1.411
Differences between Readings	60	1.375
For deviations of means, $F = 4.14/1.411 = 2.94$.	$F_{.01} = 2.45$	

It is evident that the vitamin concentration of the packages deviated significantly from the regression surface. This raises the question as to whether the empirical equation is adequate to describe the population regression. The available evidence is displayed in Table 3.

TABLE 3

Deviations of Treatment Means From Regression Surface

Storage Time (Weeks)	Storage Temperature (Fahrenheit) 0°	5°	10°	20°
0	+0.7			
2	−0.5	−1.1	+0.2	−1.4
4	+0.2	+1.0	+0.4	−0.8
6	−0.1	−0.3	+0.7	−0.3
8	0.0	−1.2	+0.2	+0.8

A deviation of 1 unit is significant at the 5% point

Four of the 17 points deviated more than the significant difference, three of them being at the 5° storage temperature. This suggested that a more flexible surface might be required; consequently, the equation

$$Y = \alpha + \beta S + \gamma S^2 + \delta ST + \eta S^2 T^2 + e, \tag{3}$$

was fitted. This allowed parabolic curves in both directions, which seemed to be indicated in some similar experiments on other products. The resulting fit was not so good as that of the simpler equation: $R^2 = 0{\cdot}8621$. Doubtless cubics would have the fit; but the propriety of using such elaborate functions with limited data is doubtful. Since the distribution of positive and negative deviations in Table 3 is fairly unsystematic, it was decided to accept the simple model tentatively and to explore other possibilities of explaining the excess of large deviations.

The only flaw detectable in the design was this: between eight o'clock and noon, some of the snapbeans may have lost ascorbic acid while standing in the uncooled room. The remedy would be to prepare the packages and send them to the quick-freeze chamber in batches or blocks, each batch containing one package per treatment. This randomized blocks design would insure for all treatments one package from each processing period: if there is any loss due to standing, it will be equalized among the treatments and will also be eliminated from error. This modification of the design is being tried. Perhaps it will be learned whether this explains the significant deviations of the treatment means from regression.

The other information concerning improvement of the design came from the analysis of variance of the two determinations of ascorbic acid which were made on each package. Here is made the assumption that the nonrandom deviations from regression have disappeared so that the package variation within treatments is the appropriate estimate of error. The mean square for differences between readings (1·375, with standard error 0·247) may be taken as an estimate of the variance, σ^2, of a normal deviate. The mean square for packages then becomes $2\sigma_p^2 + \sigma^2$, where σ_p^2 is the component of variance associated with packages. Thus σ_p^2 is estimated by

$$\frac{1.411-1.375}{2} = 0.018,$$

with standard error 0·193.

With estimates of these two variances, the potential efficiencies of similar experiments may be examined under various cost situations. If P stands for the number of packages and R for the number of readings on each in the redesigned experiment then the estimated amount of information is

$$\text{(4)} \qquad \frac{1}{s^{2x}} = \frac{PR}{1.375 + 0.018R}$$

Furthermore, representing by C the cost of a reading relative to unit cost of a package, the cost of evaluating a treatment mean will be

$$\text{(5)} \qquad P(1 + RC)$$

The maximum amount of information for a specified cost may be shown to correspond with

$$\text{(6)} \qquad R = \sqrt{\frac{1.375}{0.018C}} \text{ and } P = \frac{C_0}{1+RC},$$

where C_0, the predetermined cost of a treatment mean, is here taken as the original expenditure in this experiment, $3(1 + 2C)$.

Usually P, as determined from (6), will be a fraction: then the nearest integer is to be chosen as the optimum value, P'. The corresponding value of R is

$$\text{(7)} \qquad R' = \frac{C_0 - P'}{P'C}$$

The associated optimal amount of information will not be much different from the maximum. As compared with the original, the newly designed experiment may be expected to yield more information for the same cost.

As illustration, examine first the case in which the cost of an asc acid determination, C, is one-fourth the cost of materials, labor storage of a package.

Then, since $\sqrt{1.375/0.018} = 8.74$ and $1/C = 4$,

$$R = 8.74\sqrt{4} = 17.48,\ C_0 = 3(1+2/4) = 4{\cdot}5, \text{ and } P = \frac{4{\cdot}5}{1+17.48(1/4)} = 0.84$$

Taking $P' = 1$, the optional number of readings per package is

$$R' = \frac{4.5-1}{1(1/4)} = 14$$

At the same expenditure of funds, the redesigned experiment may be expected to furnish (equation 4)

$$\frac{(1)\ (14)}{1.375 + (0.018)(14)} = 8.605$$

units of information instead of the original 4.252 units, a relative efficiency of 202%. This may be compared to the maximum relative efficiency, 204%.

If the cost of materials and storage is not charged to the experiment, as is usually the case in experimental work, the cost of chemical analysis and packages may be about equal; whence $C = 1$. This assumption leads to $P' = 1$, $R' = 8$, and relative efficiency equal to 124%.

The experiment as conducted was approximately 100% efficient at the cost $C = 19$.

Among other vegetables investigated, broccoli offers some interesting contrasts with snapbeans. The mean concentrations of ascorbic acid in the several treatments, Table 4, show a tendency, at the lower temperatures, first to decrease then to increase with length of storage.

TABLE 4

Average Concentration of Ascorbic Acid in Broccoli

(mg. per 100 g.)

Storage Time	Storage Temperature (Fahrenheit)			
(Weeks)	0°	5°	10°	20°
0	68.3			
2	67.9	65.0	61.0	55.3
4	64.4	61.9	59.7	47.0
6	59.2	52.3	54.3	39.6
8	69.1	64.7	60.5	37.5

The explanation of this is not known. There are chemical interactions among the ascorbic acid and related compounds, but no reasonable explanation of this peculiarity has been suggested.

In this circumstance, one would usually retain the reasonable equation used to fit the snapbean data. But since the objective here is to study the structure of the experiment, a more flexible surface was adopted, the curves being parabolas in both directions. The fitted equation was

$$Y = 69.20 - 3.701S + 0.4650S^2 - 0.1259ST - 0.000481S^2T^2, \tag{10}$$

with $R^2 = 0.7446$. The analysis of variance is shown in Table 5.

TABLE 5

Analysis of Variance of Ascorbic Acid Concentration in Broccoli

Source of Variation	Degrees of Freedom	Mean Square	Estimation of Variance Components
Regression	4	2,256	
Deviations of treatment means	12	81.48	
Packages	42	50.38	$s_p^2 = 24.678$
Readings	59	1.032	$s^2 = 1.032$

While the deviations of treatment from regression are not significant, the one at $S = 6$ weeks, $T = 5°$ is -7.2 units, almost twice as great as the largest deviation. This deviation was attributable to one package whose two readings were 48.9 and 39.4 compared to the average reading 60.2 for packages having approximately the same regression value. Apparently, something happened to this package; it is doubtful if the new design of randomized blocks will account for so extraneous a value.

From the components of variance it is clear that with any ordinary cost relations only one reading need be taken on a package. Only if the cost of a package were 96 times the cost of a reading would the maximum information be got by taking two readings per package.

With one reading per package, the number of packages with constant total cost is

$$P = \frac{C_0}{1 + C} \tag{11}$$

It is easily found that, as two examples:

(1) If $C = 0.25$; $P' = 4$, $I/I_0 = 131\%$

(2) If $C = 1$; $P' = 5$, $I/I_0 = 163\%$

There is no intention of intimating that a redesigned experiment will produce just the information predicted from the one in hand. Even after repeated trials there will be the inevitable sampling variation. Besides, the effect of uncontrolled environmental conditions, such as weather, is usually not adequately known. To make allowance for these contingencies it is wise to provide a moderate flexibility in the design. In the experiment on snapbeans, at least two packages could probably be retained for each treatment under any reasonable cost situation.

Another great advantage of flexibility is this: some modification in design is often feasible after the preliminary data are available for examination. In the case of snapbeans, for example, one set of packages is assayed for the vitamin immediately after quick-freezing, providing new and more intimate estimates not only of σ^2 and σ_p^2, but also of the costs. Based on this knowledge, the number of readings on the stored packages may readily be adjusted to increase prospective information.

It may be in order to place greater emphasis on this utilization of statistics during the progress of an experiment, especially one that extends through several phases or seasons. Experimenters are prone to leave until the last minute the more revealing parts of their statistical analyses. When the experiment is completed and the summaries are finally made, it often becomes obvious that much of the potential information was lost through lack of an efficient design. This could usually be avoided by thoroughgoing examination of the data at each stage of the experiment. In some instances, complete redesign or even abandonment may be indicated. Probably in most cases modifications will increase efficiency.

The experiments discussed above are quite inflexible because all the material had to be prepared and stored in advance. But it may be worth noting that the data on packages which were not stored could have been analyzed statistically during the first week. The results in Table 6 would have been available:

TABLE 6

Analysis of Variance of Packages Quick-Frozen but Not Stored

Source of Variation	Degrees of Freedom	Mean Square	Estimates of Components of Variance
		Snapbeans	
Packages	7	0.833	0.
Readings	8	2.047	2.047
		Broccoli	
Packages	10	71.622	35.465
Readings	11	0.692	0.692

For both products, the indicated changes in design are substantially the same as those revealed at the end of the experiment. But two queti different courses of action would be indicated. For the snapbeans, there would be no point in discarding any of the packages because their cost has already been incurred; that is P is fixed at 3; and therefore, for no change in total cost, $R = 2$. But it would have been clear that information could be increased in proportion to the number of readings per package. If readings were cheap or if increased information were highly desirable additional readings might have been decided upon.

In the case of broccoli, the extra reading on each package was clearly not very productive of information. If readings were expensive, it might have been decided to save half their cost at the sacrifice of a small amount of information. For $P = 3$ and $R = 1$, the ultimate amount of information in this experiment would have been decreased only 2% at a saving of $3C$ in the total cost per treatment. For $C = 1$, this is a saving of one third on each of the remaining treatments.

The second illustration presented in this paper is drawn from a similar kind of investigation, one in which the keeping quality of snapbeans was tried with five storage-temperature treatments. The storage periods were zero, three, and six months: during the latter two, there was maintained a temperature of 0°F together with an oscillating temperature between 0° and 10°. The oscillation was intended to simulate the conditions in a home freezer. For each treatment three packages of the beans were prepared and frozen on each of four successive days, that is, in four lots. The variable measured was flavor, assayed organoleptically by four judges.

It was clear to the designers that one of two effects would be confounded. If some of the beans were held in storage before the treatments were applied so that all storage would end at the same time, the effects of the preliminary holding would be unknown. On the other hand if the treatments were to begin coincidentally, assays would be made at three month intervals and at each period the judges would be aware of the storage treatment which had been applied. Since the interest was in the contrast between steady and oscillating temperatures, the latter horn of the dilemma was chosen.

Looking at the experiment as a whole, the analysis of variance in Table 7 shows that the judges did not agree as to the differences among treatments: the judge × treatment interaction is large relative to $JTL = 6.56$. Clearly, if the judges are assumed to be a random assortment, this discrepancy is a real source of error in testing treatments, and should be combined with the randomized blocks discrepance, $TL = 8.20$.

TABLE 7

Analysis of Variance of Scores of Four Judges on the Flavor of Snapbeans Kept Under Five Storage-Temperature Treatments.[a]

Source of Variation	Degrees of Freedom	Mean Square	Estimates Furnished by Mean Square
Treatment	4	34.61	$\sigma^2+3\sigma^2_{JTL}+4\sigma^2_{TLP}+12\sigma^2_{JT}+12\sigma^2_{TL}+48\sigma^2_{T}$
Lot (Day)	3	6.30	
TL	12	8.20	$\sigma^2+3\sigma^2_{JTL}+4\sigma^2_{TLP}+12\sigma^2_{TL}$
Packages	40	3.52	$\sigma^2+4\sigma^2_{TLP}$
Judge	4	65.90	
JT	12	38.16	$\sigma^2+3\sigma^2_{JTL}+12\sigma^2_{JT}$
JL	9	9.19	
JTL	36	6.56	$\sigma^2+3\sigma^2_{JTL}$
JP	120	2.64	σ^2

$s^2 = 2.64,\ s^2_{JLT} = 1.307,\ s^2_{JT} = 2.633,\ s^2_{TLP} = 0.220,\ s^2_{TL} = 0.063$

The mean square with component,

$$\sigma^2 + 3\sigma^2_{JTL} + 4\sigma^2_{TLP} + 12\sigma^2_{TJ} + 12\sigma^2_{TL},$$

can be estimated by

$$TL + JT - JTL = 8.20 + 38.16 - 6.56 = 39.80,$$

and since this is greater than the treatment mean square, the treatment differences may be only sampling variation.

In other experiments of this type, a test of the significance of treatment effects is needed. An approximate test was suggested by Daniels.[1] Since three mean squares with unity coefficients are involved in the estimate of error, the *F*-test may be used with degrees of freedom,

$$\frac{(TL+JT-JTL)^2}{\frac{(TL)^2}{12}+\frac{(JT)^2}{12}+\frac{(JTL)^2}{36}} = 12$$

This gives $F = 34.61/39.80 = 0.87$ with degrees of freedom 4 and 12 for the approximate test.

It became evident from both this experiment and the one on ascorbic acid that the nutritive value and the flavor of snapbeans were unaffected by storage at the constant temperature of 0° F. This indicates a method of avoiding the confounding pointed out above. It seems that the beginning of the storage periods with higher temperatures can be delayed so that all the treatments can be completed at the same time without impairing the estimates of vitamin concentration and flavor. The variously treated products can then be presented to the judges in unknown and random order, thus eliminating from the treatment means the effects of possible prejudices on the part of the judges.

---

[a] Dr. A. M. Mood kindly furnished the components of variance.

[1] *Supplement to the Journal of the Royal Statistical Society,* 6:186 (1939). See, also, Satterthwaite: *Psychometrika,* 6;309 (1941); and *Biometrics Bulletin,* 2:110 (1946).

Whether this will decrease the judge × treatment interaction remains to be seen. Observation of many experiments raises a strong doubt that it will. The hazard of disagreement among judges lowers the precision of most such trials: the coefficient of variation in this one was 31.4%.

As before, the efficiency of allocation of experimental facilities may be examined. The information per treatment mean is

$$\frac{PLJ}{s^2 + Ps^2_{\mathrm{JTL}} + Js^2_{\mathrm{TLP}} + PLs^2_{\mathrm{TJ}} + PJs^2_{\mathrm{TL}}}$$

while the cost of a treatment mean is

$$PL(1 + JC),$$

where C is the cost of a judgment relative to unit cost of a package.

Examination of the information function leads to the conclusion that, for any chosen values of J and of PL, the information might have been increased by decreasing P (unless $\sigma_{\mathrm{TL}} = \sigma_{\mathrm{JTL}} = 0$). This fact was available before the experiment was performed. Had the experimenters been aware of it, they probably would have made $P = 1$ and $L = 12$ with what turned out to be an increase of 10% in information with no increase in cost. So far as *statistical* efficiency is concerned, a single individual per plot always wins.

Substituting $P = 1$ together with the values of the variance components from Table 7, information becomes

$$\frac{LJ}{3.947 + 2.633L + 0.283J}$$

It is now clear that for any fixed number of judgments, LJ, information is maximized by minimizing the quantity,

$$2.633L + 0.283J,$$

which is done by taking $J/L = 2.633/0.283$ ($= 9.3$ approximately). Evidently the original design could be improved by increasing the number of judges at the expense of lots of packages. Furthermore, the cost would be decreased, because, for any constant product, LJ, the cost, $L + LJC$, decreases with L. This means that, in the experiment producing maximum information at a fixed cost, $J/L \geqslant 9.3$. It turns out that in this optimum experiment, for values of C in the neighborhood of unity, J/L is little greater than 9.3. As examples,

$$\text{if } C = 1, \quad J = 23.9, \ L = 2.41, \ J/L = 9.9$$

$$\text{if } C = 0.25, \ J = 30.7, \ L = 2.76, \ J/L = 11.1.$$

The necessity of choosing integral values for L and J lessens the importance of the distinction between the optimum value of J/L and the value $J/L = 9.3$. Practically, the optimum experiment may be specified by $L = 3$ and $J = 28$. The information to be expected from this redesigned experiment is 352% of that in the original.

The inefficiency of the original design resulted chiefly from the large component of variance, $s^2_{JT} = 2.633$. Should this be typical (there is some evidence that it is not unusual), the fact must be faced that information gained by organoleptic methods is expensive. Since large numbers of judges are not feasible, increasing information can be got only at the expense of expanding the number of replications. Each increment of information is gained at the expense of an additional number of dollars per unit.

Résumé

Les résultats d'une expérience bien menée d'après des plans bien préparés, renseignent non seulement sur le sujet étudié, mais encore sur les améliorations à apporter au plan en vue d'expériences semblables ultérieures. Après avoir retiré de l'expérience les résultats immédiats qu'il désire, l'expérimentateur est tenté d'abandonner les données, et de procéder à une autre expérience sans tirer profit des connaissances qui auraient pu être obtenues en ce qui concerne l'amélioration du plan. Certains même, vont jusqu'à répéter la même expérience plusieurs fois de suite pour finir par arriver à la simple constatation que le plan était inefficient ou incorrect. Cela signifie un gaspillage de temps et d'argent. Dans la présente étude, l'auteur donne une illustration de la méthode d'analyse de deux expériences dans lesquelles le plan apparaît être susceptible d'amélioration.

Une expérience "factorielle" a été faite pour déterminer la diminution du contenu d'acide ascorbique de haricots verts conservés, pendant 4 périodes (2, 4, 6, et 8 semaines), à chacune des 4 températures: 0, 5, 10, et 20 degrés F. Trois paquets de haricots furent soumis à chaque traitement. Pour chaque paquet on détermina, par deux mesures, la concentration de la vitamine. Leurs moyennes sont données dans le Tableau 1.

TABLEAU 1

Concentration moyenne d'acide ascorbique dans les haricots verts
(mgr. par 100 gr.)

Période de conservation (Semaines)	Température pendant la conservation (Fahrenheit) 0°	5°	10°	20°
0	16.2			
2	15.0	14.2	14.9	11.4
4	15.7	16.0	14.3	9.2
6	15.4	14.4	13.7	7.0
8	15.5	13.2	12.3	5.3

L'équation $Y = \alpha + \beta ST + \gamma ST^2 + e$ a été ajustée aux observations, où Y, S, et T représentent respectivement la concentration d'acide ascorbique, la durée de la conservation, et la température pendant la conservation.

TABLEAU 2

Analyse de la variance de la concentration d'acide ascorbique dans les haricots verts

Source de variation	Degrés de liberté	Carré moyen
Coefficients de régression	2	531.2
Ecarts des moyennes par rapport à la régression	14	4.14
Déviation des paquets par rapport aux moyennes	43	1.411
Différences entre les mesures	60	1.375

La moyenne par paquet présentant une déviation significative de la surface de régression, il faut constater que la méthode des blocs répartis au hasard pourrait être adéquate. Il y a alors pour chaque traitement un seul paquet provenant d'une période de préparation.

La valeur informative a été examinée sous différents rapports entre les frais de mesure de la concentration d'acide ascorbique et le coût d'un paquet. L'expérience telle qu'elle a été menée offrait un résultat optimum seulement si les frais d'une mesure étaient 19 fois le coût d'un paquet. Les mesures étant relativement moins chères que le montant correspondant à ce rapport, un plus grand nombre de mesures par paquet et un plus petit nombre de paquets aurait procuré un montant d'information plus élevé par unité de frais.

Dans une expérience similaire, la saveur et d'autres qualités du produit ont été jugées organoleptiquement par quatre juges. La conclusion principale était la suivante: les juges étaient en si grand désaccord dans leurs jugements de la qualité, que l'information par unité de frais peut être rendue maximum seulement si l'on augmente sensiblement le nombre de juges.

Discussion

Mr. **Yates** gave a few examples of the value of examining experimental results to improve accuracy in future experiments on similar material. In the first, he wished to decide how far "compounding" was worthwhile. By use of experimental data, he had learned how valuable the Latin square is, relative to randomized blocks. In the second, Mr. Yates wished to discover whether rotation of treatments, eliminating animal differences, produced more efficient results than constant treatments, eliminating carryover effects of treatments.

Mr. **Sukhatme** noted that research in the theory of experimental design has been confined to modifications of factorial design, by devices like compounding, fractional replication, etc. He believes that, since factorial experiments are practicable only at large experiment stations, there should be more appreciation of the utility of small experiments, for purposes of measuring the effect of improvements introduced on the basis of results from large experiments. Mr. Sukhatme described a sampling plan which has been in use in India for such experiments. An important factor was that the design in each farmer's field had to be kept very simple.

LIMITATIONS AND POTENTIALITIES OF THE APPLICATION OF THE THEORY OF EXPERIMENTAL DESIGN IN SOCIOLOGY

by Margaret Jarman Hagood
Bureau of Agricultural Economics, Department of Agriculture
(United States)

The theory of experimental design, as developed by Professor R. A. Fisher, has had far less application in the social sciences than in the biological sciences; in fact, examples are almost nonexistent. Therefore, it seems appropriate first to examine the reasons for the limitations, so as to reveal the areas in which the theory does have potentialities of exploration and also to reveal wider areas in which the Fisher methods of analysis may be applied to data not obtained through experiment.

SOCIOLOGY NOT PRIMARILY AN EXPERIMENTAL SCIENCE

The phenomena involved in the central problems of sociology, as in the case of economics and political science, are relatively nonmanipulable for research purposes. Experimentation in the sense that a given treatment or combination of treatments may be applied to observational units with specified characteristics and the results recorded for statistical analysis is not generally feasible.[1] For example, the sociologist cannot take action to initiate disruptive forces in communities or groups he may be studying, even though he may be able to conceive of an experiment that would test the efficacy of various measures for combatting disruptive forces. Similarly, an economist who has a hypothesis as to the type of fiscal policy that would bring about quick recovery from a depression cannot decree a series of depressions in which alternative policies could be tested in order to determine experimentally the validity of his hypothesis. Some fifty years ago an editor of Sir William Petty's and Captain John Graunt's economic and population writings ascribed the development of Political Arithmetic, "a beginning of what is now called statistics," to the very fact that the experimental method is precluded in economic investigations.[2]

---

[1] Sociologists are by no means in agreement in usage of the term "experiment" in sociological research. For a review of current conceptions on the nature of the experimental method, see Ernest Greenwood, *Experimental Sociology: A Study in Method,* New York, King's Crown Press, 1945, pp. 7-18.

[2] *The Economic Writings of Sir William Petty together with the Observations upon the Bills of Mortality More Probably by Captain John Graunt,* edited by Charles Henry Hull, Cambridge, 1899, p. lxv.

In the areas of social research where experiment is impossible, at least for research workers, it seems necessary to grant that the principles of experimental design cannot be directly applied. This does not mean, however, that the methods of Professor R. A. Fisher have no potential contribution to social research.

AREAS IN WHICH EXPERIMENTAL DESIGN IS APPLICABLE

In certain research areas of sociology which deal with individuals or small groups and with "innocuous" treatments, experiments can be carried on and the theory of experimental design can be applied. For examples of application of the Fisher principles of design of experiment, however, it is necessary to go outside the field of sociology to closely related disciplines. I have been unable to find any published account by a United States sociologist of a *sociological* experiment designed according to Fisher's theory.[3] In the United States, educational psychologists and statisticians have gone further than have any other psychological or social scientists (including general, animal, and social psychologists) in the use of Fisher's methods. A survey of experimental studies in education made during a recent three-year period found "approximately equal representation of studies using the earlier methods and those using the Fisher approach,"[4] although the latter classification included some nonexperimental studies that had employed Fisher's methods for analysis of the data. The experiments described in that survey included subject matter very closely related to that with which certain specialized sociologists deal and they might very well serve as suggestive guides in planning experimental studies in the field of social psychology, which sociologists as well as psychologists claim as their province.

A related area of promise for experimental method is in the development and perfecting of tools for measurement of social phenomena. For example, in the pretesting of schedules for use in social surveys, experimental design might often replace purposive or purely random methods in choice of units on which alternate versions of a schedule or scale is to be tested.

---

[3] Of the nine experimental studies described by F. Stuart Chapin in *Experimental Designs in Social Research* (New York, Harper and Brothers, 1947), none relies on Fisher's theory for *design* of experiment, although in some cases Fisher's methods of analysis are utilized. Only four of the studies reviewed involved measurement before and after some "treatment."

[4] Charles C. Peters, Agatha Townsend, and Arthur E. Traxler, Chapter IV, "Research Methods and Design," in "Methods of Research and Appraisal in Education," *Review of Educational Research*, Vol. 15 (1945), p. 377.

APPLICABILITY OF FISHER'S METHODS TO ANALYSIS OF SOCIAL AND ECONOMIC DATA

It is in the process of analysis of nonexperimentally obtained data that Fisher's methods seem to hold greatest promise for social sciences. The first published examples noted in this country appeared not quite ten years ago. In December 1937, Friedman published an article in which analysis of variance was applied to data on family expenditures.[5] Shortly thereafter an unemployment census volume was published in which Dedrick and Hansen used analysis of variance for examination of the significance of regional differences in unemployment.[6] Other published examples were infrequent until the last few years when they have become more numerous.[7]

The conditions for applicability of Fisher's methods of analysis to social and economic data are quite broad. Many types of quantitative data, gathered with sufficient identification to permit classification according to relevant criteria, may be appropriately subjected to analysis of variance and covariance.

The methods are especially appropriate when the problem is one of establishing the existence of differentials and exploring factors that are associated with differences. In its present stage of incomplete development into a science, sociology has many such problems in a wide variety of subfields. Sociological statisticians or, more correctly, sociologists with statistical competence who engage in quantitative social research, increasingly use Fisher's methods, if one judges by published works. And it is likely that in sociology, as in the case of educational psychology, the quantitative analyses made by graduate students in connection with their doctoral dissertations utilize these modern statistical methods to an even greater extent.

In summary, sociology will probably never become wholly or primarily an experimental science. But in certain areas of research that border on educational and social psychology, experimentation is feasible and there is no reason to believe that progress will not be made in the adoption of modern principles of experimental design. Of more importance to the central field of sociological research, however, will be further advances in the application of the Fisher methods of analysis to the complex systems of interrelated social and economic variables.

---

[5] Milton Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." *Journal of the American Statistical Association*, Vol. 32 (1937), pp. 675-702.

[6] Census of Partial Employment, Unemployment, and Occupations 1937, Volume IV, *The Enumerative Check Census*, by Calvert L. Dedrick and Morris H. Hansen, U. S. Govt. Printing Office, Washington, 1938.

[7] For a textbook exposition of the application of Fisher's methods to analysis of social and economic data, see Margaret Jarman Hagood, *Statistics for Sociologists*, New York, Henry Holt and Company, 1941, pp. 541-592 and 679-716.

Résumé

La théorie de plans d'éxpériences, comme elle a été développée par R. A. Fisher, a été mise en pratique beaucoup moins dans les sciences sociales que dans les sciences naturelles. Cela est du principalement à ce que les phénomènes en question ne se prêtent ordinairement pas à l'expérimentation. L'expérimentation est possible, toutefois, en certains domaines de la recherche sociologique et ceux-ci constituent un terrain potentiel pour l'application des méthodes de Fisher.

Aux Etats-Unis, les psychologues en matière d'éducation sont plus avancés que les spécialistes en d'autres domaines des sciences sociales quant à l'emploi des méthodes de Fisher. Un relevé fait récemment sur les études expérimentales en éducation montrent que, dans la moitié des cas, les méthodes de Fisher étaient employées. Puisque quelques-unes de ces études s'occupent de problèmes semblables à ceux qui concernent les sociologues, elles pourraient bien servir comme guides pour entreprendre des études expérimentales dans le domaine sociologique.

Les méthodes de Fisher ouvrent le plus de perspectives dans le domaine des sciences sociales quand il s'agit d'analyser des données non obtenues par voie d'expérimentation. La première étude aux Etats-Unis traitant de cette façon d'employer les méthodes de Fisher a été publiée il y a dix ans. Les exemples se sont multipliés récemment et l'on peut s'attendre à ce qu'ils augmentent encore.

Discussion

Mr. Mahalanobis discussed an adaptation of the Fisherian design which has been used extensively in India since 1940, for small-scale surveys.

Mr. Yates mentioned a housing survey which is underway in England, and described the design of the experiment which is to find out the reaction of the people who live in the houses to various types of housing conditions. He added a few remarks on the wartime development of the factorial design for agricultural experimentation, and on recent improvements in animal husbandry experiments.

Mr. Eisenhart referred to the work of the late Leonard Salter, consisting of a critique of research methods in the social sciences. Salter felt that the influence of Karl Pearson's *Grammar of Science* had ended to overemphasize description as against clearly stated objectives, description for its own sake in contrast to interpretation of data as a basis for action.

Mr. Baten and Miss Hagood also participated in the discussion

SOCIAL STATISTICS SECTION

Wednesday, September 10, at 4:00 p.m.

Report & Contributed Papers

CHAIRMAN :

Ph. J. Idenburg

RAPPORT SUR L'ACTIVITE DE LA COMMISSION MIXTE DE STATISTIQUE CRIMINELLE

par Alessandro A. Molinari

Directeur Général de l'Institut Central de Statistique (Italie)

1. En 1931, à Madrid à l'occasion de la vingtième session, l'Assemblée de notre Institut approuva (séance du 16 septembre) la constitution d'une Commission mixte entre la Commission Internationale Pénale et Pénitentiaire (C.P.P) de Berne et l'Institut International de Statistique (I.I.S), "pour contribuer à la préparation de recherches de statistique criminelle internationale".[1]

---

[1] A la suite d'une communication de M. J. R. B. de Roos ("Consonnes et voyelles"—Communication sur la Statistique Criminelle) à la XVIIème session du Caire (1927), l'Institut International de Statistique avait institué, dans son sein, une "Commission pour l'étude comparative des statistiques criminelles dans les divers pays." (Membres: M. F. Faure, Prof. C. Gini, Prof. M. Greenwood, Dr. J. R. B. de Roos, rapporteur, et Dr. E. Würzburger.) Le premier rapport de la Commission fut présenté par M. de Roos à la XVIIIème session deVarsovie (1929) où il fut décidé de continuer les études sur la base des recommandations et des observations du rapporteur.

L'année suivante le Xème Congrès International pénal et pénitentiaire (Prague, 1930) vota une résolution qui exprimait la possibilité et le désir de "réaliser une coopèration internationale en vue d'organiser l'étude scientifique, par des méthodes uniformes, des causes des fluctuations de la criminalité."

M. de Roos jugea alors opportun de s'adresser à la Commission Internationale pénale et pénitentiaire pour parvenir à une collaboration entre les deux institutions internationales dans le but de chercher à éviter un double emploi et d'obtenir une plus grande uniformité dans le domaine des statistiques criminelles internationales. Les démarches faites aboutirent à la création de la Commission mixte mentionnée dans le texte.

Depuis lors la Commission mixte s'est réunie trois fois: à Leipzig (24 novembre 1932), à Paris (27 et 28 janvier 1936), à Berlin (18 et 19 avril 1939). La guerre a interrompu les travaux qu'il s'agirait maintenant de continuer. Avant de prendre une décision à ce sujet il paraît nécessaire de récapituler le travail accompli par la Commission.

2. Dans la première réunion de Leipzig[2] la Commission mixte, après avoir énuméré les difficultés qui font obstacle à une comparaison internationale des statistiques criminelles, fixa comme suit son programme de travail:

a) Préparation d'un schéma général uniforme d'"introduction"—dans une langue d'usage international—qui devrait accompagner les publications nationales de statistique criminelle afin de les rendre compréhensibles aux personnes intéressées de tous les pays.

b) Préparation d'un schéma des éléments qui devraient, de préférence, être rassemblés pour dresser les statistiques criminelles de chaque pays et le mode de leur publication. Dans ce but il est nécessaire de prendre connaissance des formulaires utilisés dans les divers pays pour recueillir les données fondamentales.

c) Etudes concernant la possibilité d'instituer une enquête internationale dans un domaine spécial nettement déterminé, du droit positif. Dans ce but la Commission internationale pénale et pénitentiaire devrait arrêter une définition internationale de l'infraction destinée à former l'objet de l'enquête.[3] (On envisagea: les infractions contre les moeurs dans lesquelles l'instinct sexuel joue un rôle; les délits de faillite ou de banqueroute; les crimes de fausse monnaie). Cette méthode implique pour une grande partie des cas, la consultation des pièces des procès afin de pouvoir constater si l'acte punissable tombe sous la définition de l'infraction telle qu'elle a été arrêtée par la Commission mixte.

M. E. Schäfer fut chargé de préparer un projet au sujet de deux questions envisagées aux points (a) et (b). M. le Comte Gleispach se chargea d'élaborer des projets en vue de l'enquête envisagée au point (c).

---

[2] Ont participé à cette réunion, comme représentants de la C.P.P.: M. le Comte W. Gleispach et le Dr. E. Schäfer; comme représentants de l'I.I.S.: MM. le Prof. G. Mortara, le Dr. J. R. B. de Roos, et le Dr. Würzburger. Etaient également présents: MM. N.L.J. van Buttingha Wichers, sous-Directeur de l'Office Permanent de l'I.I.S. et le Dr. E. Roesner du *Statistisches Reichsamt.*

[3] Cette idée se rattache aux propositions des MM. C. Gini et A. Spallanzani présentées à la session de Madrid en 1931 dans la communication ayant pour titre "*Sulla comparazione dei dati di statistica criminale dei diversi Stati*" (Sur la comparabilité des données de statistique criminelle des divers Etats). Les auteurs proposaient d'établir *ex novo* la définition juridique d'un certain nombre de délits de façon à pouvoir en déduire, avec des méthodes statistiques uniformes, des séries de contenu identique.

De la discussion il ressortit clairement que le programme préparé par la Commission mixte visant les points (a) et (b) ne constituait qu'un premier pas dans la voie d'une comparabilité internationale, c'est-à-dire un travail *préliminaire* indispensable pour le développement ultérieur des études. Dans cette première phase du travail, la Commission mixte a écarté intentionnellement la question de la comparabilité internationale des statistiques criminelles telles qu'elles viennent d'être publiées par les divers pays, non seulement en ce qui concerne l'*état* de la criminalité—dont la comparabilité a été jugée presque impossible à réaliser —mais également le *mouvement* de la criminalité.

La première tache de la Commission mixte se borne donc à réaliser une meilleure compréhension des statistiques criminelles des divers pays, et, par conséquent, à réunir tous les éléments nécessaires pour être à même de juger du degré de leur comparabilité internationale ou par groupe de pays (zones de bonne ou de mauvaise comparabilité, etc.)[4]

3. Dans la réunion de Paris (janvier 1936) la Commission mixte[5] prit en examen le projet de l'"introduction" et le "schéma-type" des cadres statistiques, préparés par M. E. Schäfer.[6]

La question concernant l'enquête internationale sur un délit déterminé fut provisoirement ajournée.

Sur la base des discussions faites à Paris et des observations successives des membres de la Commission mixte, M. Schäfer présenta, au nom de la Commission, un rapport à la XXIIIème session d'Athènes (septembre—octobre 1936) ayant pour titre: "Directives pour l'élaboration des statistiques criminelles dans divers pays."[7]

L'Assemblée de l'Institut (séance plénière du ler octobre) adopta les directives recommandées par la Commission.[7]

---

[4] Les pays nordiques (Suède, Danemark, Norvège, Finlande) seraient à même d'appliquer les recommandations adoptées par l'Institut en 1936. Cela rèsulte de l'étude publiée dans la *Revue* de l'Institut International de Statistique (11e année, livraison 3/4, 1943: "La statistique criminelle et pénitentiaire des pays nordiques," par le Dr. E. J. Höijer, directeur en chef du Bureau central de Statistique, Stockholm, et par le Dr. Otto Grönlund, Chef du Bureau, Stockholm).

[5] Ont participé à cette réunion comme représentants de la C.P.P.: MM. A. Mossé, H. Nissen; le Dr. E. Schäfer; comme représentants de l'I.I.S.: MM. le Dr. A. Molinari, le Dr. H. Truchy, le Dr. E. Würzburger. Etaient également présents: le Prof. J. Simon van der Aa, secrétaire général de la C.P.P. et M. N. L. J. van Buttingha Wichers, sous-directeur de l'Office Permanent de l'I.I.S.

[6] En 1934 et 1935 les membres de la Commission mixte avaient examiné le premier projet du Mémoire préparé par le Dr. Schäfer et lui avaient fait parvenir leurs observations qui furent, en même temps, portées à la connaissance de tous les membres de la Commission mixte.

[7] Le rapport a paru en français dans le *Bulletin* de l'I.I. S.—Tome XXIX 3ème livraison (pages 240 à 251). Dans le même *Bulletin*, Tome XXIX, lère livraison, ont été reproduits (en anglais, en français, et en allemand): le plan pour une introduction générale à la statistique criminelle et le schéma-type des cadres statistiques (pages .5)à227

4. A cause de la guerre les décisions approuvées par la Commission mixte et par les deux organisations internationales intéressées, n'ont pu avoir une suite pratique. L'Office Permanent de l'I.I.S devrait, par conséquent, s'adresser de nouveau—si possible avant la fin de 1947—aux bureaux de statistique des divers pays qui s'occupent des statistique criminelles, pour les prier: (a) de faire précéder les publications de statistique criminelle d'une "Indroduction" conforme à celle qui a été approuvée par la C.P.P. et l'I.I.S.; (b) de dépouiller les statistiques de la criminalité suivant le schéma-type des cadres statistiques, également approuvé par la C.P.P. et l'I.I.S. ou, tout au moins, de faire en sorte que les statistiques nationales puissent être facilement ramenées aux cadres statistiques dont il est question.

5. A la suite des accords qui ont eu, lieu en 1938 entre la C.P.P. et l'I.I.S., il fut décidé de traiter à la réunion de Berlin[8] les questions suivantes:

a) statistique internationale des meurtres

b) statistique internationale pénitentiaire

La question sous (a) rentrait dans le programme de travail de la Commission mixte (c'est-à-dire "l'institution d'enquêtes internationales dans un domaine spécial nettement déterminé du droit positif": voir paragraphe 2) qui avait été provisoirement ajourné à Paris. Bien que le meurtre ne figurait pas parmi les infractions qui, d'après les décisions prises à Leipzig, devraient être choisies pour l'enquête internationale envisagée, il ne fait aucun doute que l'homicide est un des délits sur lequel d'éminents statisticiens et criminalistes ont insisté davantage, pendant trois quarts de siècle, pour en faire l'objet principal de statistiques internationales comparables, tout au moins pour ce qui concerne le "mouvement" dans le cours du temps.

La question sous (b) suggérée par la C.P.P., représente une addition importante au programme original de la Commission mixte.

6. Pour ce qui concerne la *statistique internationale des meurtres*, une étude très approfondie du Dr. Roesner sur la statistique internationale des homicides fut soumise à la Commission.[9] Cette étude renfermait

---

[8] Ont participé à cette réunion, comme représentants de la C.P.P.: MM. le Dr. Nissen et le Dr. Schäfer; comme représentants de l'I.I.S.: MM. A. Molinari et le Dr. Reichardt, président du *Statistisches Reichsamt*. Etaient également présents, et ont pris une part active à la discussion: MM. le Dr. Roesner, du *Statistisches Reichsamt*; Ernest Delaquis, secrétaire général de la C.P.P., et van Buttingha Wichers, sous-Directeur de l'Office Permanent de l'I.I.S. (avril 1939)

[9] Dr. Ernst Roesner, "*Internationale Mordstatistik —Vorschlage zur Methode und Technik*" (Statistique internationale des homicides—Propositions de méthode et de technique). *Monatsschrift fur Kriminalbiologie und Strafrechts Reform* (Revue mensuelle de biologie criminelle et de la réforme du droit pénal) 1939, *Heft* 2, J. F. Lehmanns, München-Berlin)

des propositions pratiques sur la méthode à suivre pour le relevé et l'enregistrement des données statistiques, leur dépouillement et leur classification. Il fut également soumis aux membres de la Commission mixte, un projet de formulaire, très détaillé, pour l'enquête statistique spéciale sur le meurtre.

En raison du manque de précision des notions juridiques au sujet de l'homicide, la Commission, unanime, proposa, d'abord, de limiter l'objet de l'enquête internationale aux délits d'homicide intentionnels (à l'exception de l'infanticide).

Le plan de travail envisagé par la Commission peut être résumé comme suit:

a) le Bureau permanent de la C.P.P. demandera aux autorités judiciaires des divers pays un relevé des dispositions en vigueur sur les délits d'homicide intentionnel et un résumé statistique des condamnés par jugement définitif au cours des cinq dernières années.

b) la Commission mixte, en tenant compte des résultats de l'enquête mentionnée sous (a) fixera la notion de "délits d'homicide intentionnels."

c) ensuite l'Office Permanent de l'I.I.S. recommandera aux divers pays d'effectuer des enquêtes statistiques sur les délits d'homicide intentionnels (selon la définition mentionnée sous (b)), en utilisant à cet effet les actes judiciaires. Les données et les informations à fournir seront enregistrées dans un formulaire dont le schéma définitif devra être préparé par la Commission mixte.

d) les résultats de l'enquête seront examinés et dépouillés par la Commission mixte.

Si la vingt-cinquième session de notre Institut est d'accord on pourra déjà passer sans autre délai à l'exécution de la proposition indiquée sous (a).

Le texte des décisions de la Commission mixte est reproduit dans l'*Annexe*.

7. A l'égard de la statistique *internationale pénitentiaire* le Dr. Reichardt soumit à la Commission mixte un rapport et des documents qui formèrent la base des discussions.[10]

---

[10] Le rapport du Dr. Reichardt, "Observations et suggestions pour la résolution du probleme de la statistique internationale des Prisons" (*Bemerkungen und Anregungen zur Losung des Problems der Internationalen Gefangnisstatistik*) était complété par trois annexes:.

1) bibliographie des congrès internationaux qui ont traité des questions de statistique pénitentiaire;

2) bibliographie sur le problème de la statistique internationale des prisons:

3) projet de questionnaire proposé par M. D'Orelli au Congrès Pénitentiaire International de Stockholm en 1878.

Furent aussi distribués deux mémoires du Dr. Roesner (publiés en 1937): "*Bibliographie der Strafvollzugsstatistik des In- und Auslandes*" et "*Die Gefangenen der Welt in Zahlen.*" Les données qui figuraient dans cette dernière étude ont été tirées, pour la plupart de la publication "*The Prisoner Population of the World*" (Septembre 1936) présentée par la *Howard League of Penal Reform*, Les résolutions adoptées par la Commission mixte sont reproduites dans l'*Annexe*.

8. La réunion de Berlin a déblayé le terrain d'un grand nombre de questions d'ordre général soulevées par les deux problèmes envisagés. La Commission mixte non seulement est tombée d'accord sur la résolution de questions statistiques fondamentales, mais elle a aussi indiqué clairement la voie pratique à suivre pour atteindre des résultats concrets et conclusifs. Toutefois il reste un grand travail à accomplir.

9. Par suite de la perte douloureuse de la plupart des éminents collègues qui faisaient partie de la Commission mixte, il serait nécessaire de la reconstituer, de sorte que leurs travaux, si fructueux et intéressants, puissent continuer. A mon avis la nouvelle Commission devrait borner son activité aux questions fondamentales et essentielles qui figurent dans le programme de travail et faire tous les efforts pour mener à bonne fin les résolutions adoptées dans les trois réunions d'avant-guerre. Un changement de programme ou, à plus forte raison, son élargissement, ne pourraient, à mon avis que nuire aux progrès, déjà si lents à réaliser, dans le domaine des statistiques internationales de la criminalité.

En conséquence je me permets de proposer que la 25ème session de l'Institut International de Statistique exprime les voeux suivants:

1) que la Commission mixte de statistique criminelle soit reconstituée et continue ses travaux sur la base des résolutions adoptées dans les trois réunions d'avant guerre, tout en tenant compte des modifications éventuelles suggérées par la nouvelle situation de l'organisation statistique internationale.

2) que les Offices Permanents de la Commission Internationale Pénale et Pénitentiaire et de l'Institut Internationale de Statistique, donnent exécution au plus tôt et autant que possible, aux enquêtes de leur ressort, suivant les décisions adoptées avant la guerre. (Voir paragraphe 4 et point (a) du paragraphe 6 du présent rapport).

Annexe

Résolutions adopteés par la Commission mixte pour la Statistique criminelle

(Réunion de Berlin, 18 et 19 avril, 1939)

I. *Statistique internationale des meurtres*

A. *Résolutions*:

1. Pour poursuivre la tâche qui lui a été confiée de favoriser la coopération internationale dans le domaine de la statistique criminelle, la Commission mixte estime désirable que soit entreprise une enquête

statistique internationale, au point de vue psychologique et sociologique, des délits d'homicide intentionnels.

2. La condition préliminaire du succès de l'enquête mentionnée sous (1) est une délimitation exacte de la notion "délits d'homicide intentionnels."

3. En conséquence, le Bureau permanent de la Commission internationale pénale et pénitentiaire est prié de demander aux autorités judiciaires supérieures des divers pays de lui faire parvenir le relevé des dispositions en vigueur dans leur pays sur les délits d'homicide intentionnels ainsi qu'un résumé statistique du nombre des individus condamnés par jugement définitif pour ces délits, au cours des cinq dernières années. Le Bureau permanent formulera cette requête de concert avec deux membres de la Commission mixte, a savoir MM. Schäfer et Reichardt.[11] La documentation recueillie à la suite de cette enquête sera examinée par les mêmes membres de la Commission mixte et soumise à la Commission.

B. *Recommandations*:

Pour la suite des travaux, la Commission recommande la procédure suivante:

La Commission mixte fixera très nettement, au cours d'une nouvelle réunion qui devra avoir lieu le plus tôt possible, la notion des "délits d'homicide intentionnels," notion qui sera déterminante pour l'enquête statistique mentionnée sous A/1.

Ensuite, l'Office Permanent de l'Institut International de Statistique recommandera aux divers pays d'effectuer, pour des périodes déterminées, des enquêtes statistiques sur les délits d'homicide intentionnels dans le sens ci-dessus, en utilisant les actes judiciaires se rapportant à des cas jugés définitivement, selon le schéma qui figure dans le formulaire annexé au procès-verbal auquel la Commission mixte se réserve de donner une rédaction définitive lors de sa prochaine réunion. Les statisticiens chargés d'établir la statistique criminelle dans les Offices de statistique de chaque pays utiliseront les actes judiciaires et rempliront les formulaires d'une manière uniforme, afin d'assurer l'homogénéité et la technicité du travail.

Les formulaires dûment remplis devront être transmis par les divers pays à la Commission mixte par l'intermédiaire de l'Office Permanent de l'Institut International de Statistique. La Commission mixte procédera à l'appréciation statistique des données contenues dans les formulaires et à la présentation numérique, sous forme de tableaux, des résultats de l'enquête.

---

[11] Malheureusement nos deux éminents collègues sont décédés.

II. *Statistique pénitentiaire internationale*

1. La Commission mixte estime désirable que, dans chaque pays, une statistique pouvant servir à la comparaison internationale soit tenue:

a) sur les *établissements* dans lesquels des peines privatives de liberté sont purgées.

b) sur les *occupants* de ces établissements, pour autant qu'ils purgent des peines privatives de liberté ou sont astreints à des mesures de sûreté comportant privation de liberté.

La Commission mixte se réserve de délimiter avec précision les établissements et les occupants entrant en ligne de compte et de désigner les conditions particulières des établissements et de leurs occupants sur lesquelles doit porter la statistique.

2. La Commission mixte résumera ses propositions sur la statistique désirée dans un schéma-type.

3. Pour la préparation de ce schéma-type, les pays seront requis de mettre à la disposition de la Commission mixte les instructions et formulaires en usage pour le rassemblement statistique des données en question. Le Bureau Permanent de la Commission Internationale Pénale et Pénitentiaire est prié d'adresser un appel dans ce but aux autorités pénitentiaires des divers pays et d'exprimer en même temps le désir de recevoir la publication la plus récente se rapportant à cette statistique.

4. Après avoir été approuvé par les deux organisations, le schéma-type devra être envoyé à tous les pays avec la recommandation d'y adapter leurs statistiques en question et de le faire précéder d'une introduction qui, sous certains points de vue, initie au domaine dont il s'agit et rende la statistique mieux compréhensible pour l'usage international.

Résumé

1. The work accomplished, from 1931 to 1939, by the joint Commission on Criminal Statistics formed by the International Penal and and Penitentiary Commission of Berne and the International Statistical Institute is summarized here.

2. *Program*

(a) Preparation of a uniform general scheme of "introduction," in an internationally spoken language, which should accompany national publications of criminal statistics to make them understood by interested persons in every country.

(b) Preparation of an outline of desirable data to be collected for com piling the criminal statistics of each country, and an outline of methods of publication.

(c) Study of the possibility of making an international investigation in some special, clearly defined, legal field. For this purpose the Commission decided to limit the international investigation to the field of murders (crimes of willful homicide, excluding infanticides).

(d) Preparation of guides for the compilation of international prison statistics.

3. *Carrying out the program.* The situation is as follows:

(a) As far as points (a) and (b) of the program are concerned, the Joint Commission has discharged its task by presenting, at the Athens Session (September-October 1936), a report that was approved by the Assembly.

(b) Concerning point (c) of the program, the Joint Commission established the theoretical and practical principles that should guide international inquiry on murders.

The Joint Commission was then to have concerned itself with establishing an exact definition of "crimes of willful homicide." To arrive at such a definition, it was recognized that it was first necessary to collect: (1) The regulations in force in the different countries concerning willful homicides; (2) a statistical summary of the number of individuals convicted by final judgment for such homicides. An investigation of this kind was to have been made by the Permanent Office of the International Penal and Penitentiary Commission, in agreement with the Joint Commission.

A first draft of subsequent tasks was outlined in the resolutions adopted during the 1939 Berlin Session (see appendix to report).

(c) With respect to point (d) of the program, the Joint Commission limited itself to making a preliminary, but intensive, study of the question. This study has enabled the Commission to agree on the objective of the investigation and on the statistical units to be considered.

(d) The Commission also prepared the agenda of future proceedings (see appendix to report).

4. In consideration of the aforesaid, the writer proposes that the twenty-fifth Session of the International Statistical Institute agree on the following draft resolutions:

(1) that the Joint Commission on Criminal Statistics be reconstituted and continue its work on the basis of the resolutions adopted in the three prewar meetings, with due regard to changes, if any, brought about by the new situation in international statistical organization.

(2) that the Permanent Offices of the International Penal and Penitentiary Commission and of the International Statistical Institute carry out, as soon and as completely as possible, the investigations in their fields which were required by the action taken before the war.

Discussion

Mr. Idenburg reported that the United Nations Secretariat had recently distributed a questionnaire concerning criminal statistics, and therefore proposed that the draft resolutions proposed by Mr. Molinari be amended to read as follows:

"(1) that the Bureau consult with the Secretary-General of the United Nations Social Affairs Department about the work undertaken by the United Nations in the field of criminal statistics.

"(2) that if after this consultation it appears useful, the Joint Commission on Criminal Statistics be reconstituted and continue its work on the basis of the resolutions adopted in the three prewar meetings, with due regard to changes brought about by the new situation in international statistical organization.

"(3) that the Permanent Office of the International Penal and Penitentiary Commission and of the International Statistical Institute carry out, as soon and as completely as possible, the investigations in their fields which were required by the action taken before the war."

The amendment was approved.

METHODS OF THE STATISTICS OF SOCIAL INSURANCE

by Demetrious Calitsounakis

Graduate School of Economic and Commercial Science, Athens (Greece)

An independent and separate system of social statistics does not yet exist. To establish such a system is very difficult for various reasons. The age group composition of a population, the great number of various economic classes into which it is divided, and moral and psychological factors prevent the objective and systematic study of a society from the statistical point of view. We are, however, going to maintain a system of social statistics as it is done everywhere else. The difficulties of such an attempt can be seen in the small progress so far achieved in the field of international comparison of labor accidents, through the joint efforts of the International Statistical Institute, of the International Labor Office, and of the previous Conferences for Social Insurance; although from the methodological point of view, many things have been clarified.

I shall try to present systematically the statistical data of social insurance in three main statistical fields. Insurance against illness and maternity belongs to the so-called biological statistics of population. Here, the statistical data should be classified according to the causes of illness. The insurance of maternity should be classified under the statistics of births.

The most important statistics of social insurance are undoubtedly the statistics of accidents. The International Statistical Institute dealt with this subject at its Tenth Session in London. The International Statistical Conference in Paris (1889) as well as that held in Berne (1891) discussed the same subject.[1] The main object of this kind of social insurance is the promotion of means for the prevention of and insurance against accidents, with the use of statistical data. Unfortunately, there are many difficulties hindering the realization of this object because international standardization depends especially on national legislation defining accidents and insurance methods. Because of this, the International Conference of Labor Statisticians (1923) dealt in general with certain definitions such as those of the nature and time of accidents, their causes, their damages, and their compensations.

Here we have to work hard and continuously to arrive at definite results. We can deal especially with labor accidents in three industrial branches, that is to say, in mining, mercantile marine, and railways.

---

[1] A good summary of the work done and of the present state of affairs is given by J. Feig, in the Proceedings of the Eighth Session of the International Statistical Institute, Warsaw (1928).

All these statistics of labor accidents[2] belong to socio-political statistics and, in a broader sense, to economic statistics.

Insurance against old age and for persons dependent on the aged, belongs to the statistics of civilization, or rather to positive moral statistics. The effect of such insurance is so beneficial that from its nature it must be treated under the above heading.

Finally, we have to deal with unemployment statistics, which belong in labor statistics or in the socio-political category. Useful comparison of the results[3] of this kind of insurance depends on the system of unemployment payments as well as on the international definition of unem ployment. We must confess that in this field great progress has been made, as can be seen in the following figures: the number of persons insured against unemployment increased from 5 millions in 1919 to 42 millions in 1934. In addition, the International Labor Office has been able to draft an index number of unemployment in the main countries.

On the above basis, I shall take the liberty of submitting the following proposal for consideration:

The permanent office of the International Statistical Institute is authorized to notify the most important countries of a classification containing systematically all fields of social statistics, and to ascertain whether they are inclined to accept this classification. In this way, it is hoped that comparable statistical data on social insurance can be obtained from the principal countries.

Résumé

Il est nécessaire de procéder à une modification systématique uniforme des statistiques des assurances sociales dans les divers pays.

Une pareille révision intéresserait surtout: (1) les statistiques de l'assurance-maladie et de la maternité; (2) les statistiques des accidents (causes, degrés, durée, indemnités); et (3) les statistiques du chômage, domaine dans lequel des progrès internationaux ont déjà été réalisés. Il serait donc utile que l'Office permanent de l'Institut International de Statistique communique aux principaux pays un rapport traitant d'une manière systématique au moins les domaines ci-dessus mentionnés de la statistique des assurances sociales et s'assure si ces pays sont disposés à l'adopter.

[2] For the classification of labor accidents see "La Standardisation Internationale des Statistiques du Travail", Bureau International du Travail, Geneve, 1934.

[3] The International Labor Office published in 1922 an essay on the methods of unemployment statistics, and in 1932 a summary of the results of the International Conferences on Labor Statistics.

Discussion

Miss **Bodmer** discussed briefly the work of the International Labor Office in the field of social insurance statistics:

"The Sixth International Conference of Labor Statisticians met in Montreal in August 1947, and adopted a resolution on the statistics of industrial accidents, including a definition of industrial accidents for statistical purposes and a standard method of computing frequency and severity rates. The subject of classification of accident causes, which at present is undergoing extensive experimentation in certain countries, has been proposed for consideration at a subsequent session of the International Conference of Labor Statisticians.

"...The Sixth International Conference dealt also with the subject of unemployment statistics and made important recommendations in this field. The *Year Book of Labor Statistics* publishes the available statistics on unemployment for the different countries of the world.

"...Under the auspices of the International Labor Office, the Medical and Statistical Technical Commissions of the Inter-American Committee on Social Security, meeting in Washington in January 1947, discussed the standardization of medical care statistics and morbidity statistics of social insurance institutions... The Joint Commissions adopted conclusions for submission to the Second session of the Inter-American Conference on Social Security, to be held at Rio de Janeiro in November 1947...

"The resolutions on the standardization of medical care statistics provide for the publication of such data as: the number of medical care service establishments—hospitals and out-patient clinics, of the medical staff employed by or working for the medical care service, the number of persons receiving care, and the items of care furnished by the service, as well as the cases and duration of hospitalization.

"As regards morbidity statistics, the resolutions adopted by the Technical Commissions of the Inter-American Social Security Committee lay down the minimum data that should be provided by social insurance institutions, such as cases of illness and cases of incapacitating illness, days of incapacitating illness, and the number of persons exposed to risk; the definition of these units, and their classification by age, sex, and disease....

"It is intended to further extend the work of the Social Insurance Section of the International Labor Office by the standardization and publication of statistics on the actuarial experience as well as the administrative aspects of social insurance institutions."

Mr. **Idenburg** suggested and participants in the meeting agreed, that Mr. **Calitsounakis** be asked to work out his ideas on the classification of statistics of society, before the International Statistical Institute should undertake work in this field.

ORIGIN, FUNCTIONS. AND FUTURE OF THE BRITISH CENTRAL STATISTICAL OFFICE

by Ernest H. Godfrey
Fellow, Royal Statistical Society (Great Britain)

The establishment of a Central Statistical Office by the British Government is an event which the International Statistical Institute will note with interest, because its publications should facilitate international comparisons, which it is the object of the Institute to promote.

What follows is an attempt to describe briefly previous efforts made by statisticians, and especially by Fellows of the Royal Statistical Society, to promote the formation of a Central Statistical Office for Great Britain, the functions which it is designed to fulfill and the lines upon which it may be expected to proceed in future.

PAST EFFORTS TOWARDS STATISTICAL REFORM

The movement that has culminated in the establishment of a British Central Statistical Office goes back for about 76 years. One of the earliest efforts towards statistical reform was made by Frederick Purdy, who was Principal of the Statistical Department of the Poor Law Board and an Honorary Secretary of the Statistical Society. In a paper read before the Society on March 21, 1871, he described numerous defects in the situation then existing, recommended the establishmen tof a department that should be intermediary between Parliament and the various Government Offices (in other words a Central Statistical Office) and laid down under eight headings the principal duties which such a department should have.

According to Sir Hubert Llewellyn Smith's book on the Board of Trade (1928), regular systematic statistics, as issued by the Board of Trade, date from 1832, when G. R. Porter, afterwards famous as author of the *Progress of Nations*, entered the Board's service. He organized the Statistical Department of the Board of Trade, and issued a kind of Statistical Year Book under the title *Tables of the Revenue, Population, Commerce, etc., of the United Kingdom and its Dependencies.* This volume appeared annually from 1832 to 1854, and from it were derived many of the later statistical publications of the Board of Trade, including various annual Statistical Abstracts. Among these may be mentioned that for the United Kingdom from 1854; for the Colonies and Possessions from 1864, later under the changed title, *Statistical Abstract for the Several Oversea Dominions and Protectorates*; for the British

Empire from 1905; for Labor Statistics of the United Kingdom from 1894; and for Foreign Labor Statistics from 1899. But the two Great Wars of 1914-1918 and 1939-1945 involved of course the permanent or temporary suspension of these publications.

Sir Robert Giffen, who was appointed head of the Statistical Department of the Board of Trade in 1876, and who acquired high reputation as a statistician, was, as stated by Sir Hubert Llewellyn Smith, a "strong advocate for centralization, not only within the Board of Trade, but also (at least in principle) within the Government Service generally. He sympathized with the idea of constituting a Central Statistical Department to serve the requirements of all the Departments of State."

On March 28, 1877, an Official Statistics Committee was appointed by the British Treasury. This Committee took evidence from numerous witnesses from March 26, 1878, to August 6, 1879, and the Minutes of the Evidence, together with a number of Appendices, were printed and published. Three Reports were presented by this Committee, the first, dated August 6, 1879, having reference to the Customs and the second, of the same date, to the statistical returns of the Local Government Authorities. The third Report, dated December 23, 1879, and published in 1881, dealt more generally with the subject under two sections: (1) principles for the revision of existing official statistics, and (2) recommendations for the future conduct and control of official statistics. They made various recommendations, including the appointment (a) of a small Central Statistical Department subordinate to the Treasury and (b) of a Board or Commission to supervise official statistics. The Treasury accepted recommendation (b), and decided to set up a Statistical Inquiry Committee.

In his presidential address to the Royal Statistical Society on November 19, 1907, Sir Charles Dilke described defects in the existing official statistics and referred favorably to proposals previously made for the establishment of a Central Statistical Office; and on June 16, 1908, Mr. Arthur Bowley, subsequently Professor of Statistics in the University of London, read a paper before the Society on the Improvement of Official Statistics and discussed the lines upon which a "Central Thinking Office of Statistics" should be set up, or a Central Statistical Department, which he had advocated on various occasions and especially in his address to a Section of the British Association in 1906.

In an article in the *Financial Review of Reviews* of January, 1911, entitled "Our Statistics: A Plea for a Central Government Department," Sir Athelstane Baines, C.S.I, also advocated the appointment of a "Central Organization, supplemented or not by a Consultative Body of experts."

On December 19, 1916, a paper entitled "The Reorganization of Official Statistics and a Central Statistical Office" was read before the Royal Statistical Society by Geoffrey Drage. This paper referred to previous efforts towards the establishment of a Central Statistical Office and concluded that a central direction of official statistics was urgently required.

In June 1919, the Council of the Royal Statistical Society appointed a Committee on Official Statistics, who prepared a Petition to His Majesty's Government representing the urgent need for a reorganization of the present system. The Petition referred to previous efforts at reform and especially to repeated suggestions which had been made for the institution of a Central Statistical Office. The petitioners finally submitted that the questions raised were of sufficient importance to justify immediate inquiry by a Royal Commission or a Parliamentary Committee. The Petition, dated November 1, 1919, was numerously signed and duly presented. To consider this Petition a Committee was appointed by the Cabinet, who reported on February 8, 1921, that in their opinion no case existed for an inquiry, whether by Royal Commission or Parliamentary Committee; and they were not in favor of the establishment of a Central Statistical Office. They recommended however the establishment of a Permanent Consultative Committee of Statistical Officers. This recommendation was adopted; and one result was the issue of an annual *Guide to Official Statistics*, which gave complete references to official publications containing statistics grouped under subjects. This proved useful; but little was attempted or accomplished in the way of improvement or coordination.

Sir Henry Rew, in his presidential address "On the Organisation of Statistics," delivered to the Royal Statistical Society on November 16, 1920, spoke as follows: "A fully equipped Central Statistical Office is desirable, competent not only to advise but charged with the responsibility of passing for publication all statistical returns and reports of sufficient importance, which are presented to Parliament or published."

CREATION OF THE PRESENT CENTRAL STATISTICAL OFFICE

So the establishment of a Central Statistical Office seemed to be as far off as ever until the conditions created by the second great world war, which broke out in 1939, accentuated the urgency of the need for such an Office.

One of the greatest changes which took place during the late war was in the collection and use of statistical information to assist in the direction of economic policy. The technique of using statistics for administrative control became very highly developed during these years, not only in preparing future programs, but also in using statistics to check the performance of each major Government program month by month or week by week. The statistical departments became concerned with providing regularly up-to-date statistical information on which programs could be made for a period ahead and by which current programs could be kept under review.

To assist in this work of using statistics intimately in day-to-day Government administration and in central control of policy, consequential changes were necessary in the organization of the Government statistical service. Central Statistical Organizations were set up in many Departments; and further, in order to ensure that the figures collected by Government Departments were prepared in a comparable and consistent manner, a Central Statistical Office was established within the Offices of the War Cabinet at the beginning of 1941, on the direction of the Prime Minister, to maintain day-to-day liaison between the Statistical Divisions of each Department and the Central Executive. The responsibility for the collection of basic information still continued to rest with the original Departments. Since the War Cabinet had to make decisions affecting more than one Department, it was essential that the statistics, for example, of imports, production, consumption, and stocks, on which these decisions must be based, should be compiled according to agreed definitions; and one of the important duties of the new Central Statistical Office was to assist in the preparation of statistics on the basis of such agreed definitions and in a form consistent with each other.

FUNCTIONS OF THE CENTRAL STATISTICAL OFFICE

The functions of the Central Statistical Office include the following: (a) as a section of the Cabinet Office it assists the Cabinet Secretariat on all questions involving the use of statistics, the preparation of reports and the collection of information needed for the Cabinet and its Committees; (b) it maintains general liaison between the Cabinet Secretariat and the Statistical Divisions of Departments; (c) it acts as a central organization for the preparation and circulation of all reports containing information obtained from more than one Department, collaborating with Departments in the preparation of such reports; and (d) it arranges for interdepartmental discussions on all statistical questions.

To carry out these functions the Central Statistical Office has built up an entirely new organization, including an administrative staff which has increased from 7 in 1941 to 21 in 1946. The Civil Estimates published annually indicate the growth of the Administrative Staff from 1941 to 1946 as follows:

Appointments	1941	1942	1943	1944	1945	1946
Director	1	1	1	1	1	1
(Chief Assistants)				4	4	
Chief Statisticians						2
(Assistants)	4	5	6	6	6	
Statisticians						8
(Junior Assistants)	2	6	10	7	6	
Assistant Statisticians						4
Chief Executive Officer						1
Senior Executive Officers						5
Total	7	12	17	18	17	21

Note: Designations in brackets became obsolete in 1946.

In addition to the Administrative Staff, as noted in the preceding table, there is a subordinate clerical staff, which according to a statement made in the House of Commons by the Chancellor of the Exchequer on November 13, 1945, amounted to 32 persons.

The duties of each Administrative Officer include (a) the maintenance of day-to-day liaison with the Statistical Divisions of Government Departments; (b) keeping himself informed of all problems of policy within his field of study as well of changing methods of compilation of the statistics with which he is concerned.

The majority of the reports and statistics prepared by the Central Statistical Office are of a confidential nature and are circulated to Cabinet meetings and Government Departments only; but the following publications have been issued by the Office: "White Papers" on War Finance and Estimates of the National Income and Expenditure for each of the years 1941 to 1946, and *Statistics Relating to the War Effort of the United Kingdom* in 1944.

From the public point of view, one of the most important of the issues of the Central Statistical Office is the *Monthly Digest of Statistics*, which was begun in January 1946. This is a large quarto of nearly 100 pages, published at the price of 2s. 6d. In the Introductory Notes to the first issue, it is stated that the Digest has been prepared by the Central Statistical Office in collaboration with the Statistics Divisions of Government Departments, and that the Departments and organizations responsible for providing each series of statistics are stated under each table. The aim is to publish in summary form statistics which show the changing pattern of the economic activity of the country. All the statistics are therefore presented in the form of time series. Figures are given, so far as possible, for each of the calendar years 1935 to 1944 and monthly since the beginning of 1944. It is stated also that the scope of the *Digest* will be revised from time to time to include new statistical information as it becomes available. Many of the series appearing in the *Digest* are given in greater detail in other official publications, as, for instance, the *Board of Trade Journal*, the *Ministry of Labor Gazette* and the *Trade Accounts*.

The Digest contains upwards of 100 Tables, comprising the following subjects: Population; Employment; Fuel and Power; Raw Materials; Manufactured Goods; Building; Agriculture and Food; External Trade; Merchant Shipping; Inland Transport; Finance and Wages and Prices. Most of the figures in the tables are expressed in thousands.

In January 1946, the Central Statistical Office issued as a supplement to the first issue of the *Monthly Digest* a sixpenny octavo pamphlet of 32 pages entitled "Definitions of Items and Units in the Monthly Digest of Statistics." This is essential even for preliminary understanding of the Tables.

LINES OF FUTURE DEVELOPMENT

In a paper "On the Future of Statistics," read before the Royal Statistical Society on April 21, 1942, Mr. M. G. Kendall, discussing the possibility of a State Statistical Office, wrote as follows: "There is now in existence an offshoot of the Cabinet Secretariat, known officially, I think, as the Central Statistical Office. Like many other English institutions it was an improvisation; like many other English institutions it may outlast the creations of the most solicitous bureaucratic tectonics." He suggested that the Royal Statistical Society should debate the question very seriously and make some recommendation to the Government.

In 1943 the Council of the Royal Statistical Society issued a "Memorandum on Official Statistics," which covered a wide field. They stated that in their view the formation of a Central Statistical Office was a development of great significance, and the principle of providing a coordinating body of this kind was one which they cordially welcomed. They set out in detail the principles that should in their view be observed in reorganizing the Office after the war; and the following are those of their recommendations which have special reference to a Central Statistical Office:

(1) Every major Department should have a statistical branch under the direction of an administrative officer; and minor Departments should either have a smaller unit or work under the guidance of the statistical branch in a major Department or of the Central Statistical Office.

(2) Firm coordination of the statistical work of different Departments is necessary. This may best be effected by a Central Statistical Office acting with a small Committee selected from among the heads of statistical branche in the major Departments.

(3) The Committee should be small, should fill its own vacancies, should be empowered to initiate subjects for discussion, and to issue reports on statistical matters not falling within the province of any one Department. Its decisions should be binding on Departments, subject, on major matters, to Cabinet authority.

(4) The Central Statistical Office should be relieved of routine analysis and administration, should, *inter alia,* undertake research work, should act generally as a coordinating body, should compile statistical information required for the Cabinet and its Committees on subjects where the interests of more than one Department are concerned and should secure prompt and full publication of Government statistics.

In May 1944, that is before the end of the war against Germany and Japan, the Minister of Reconstruction presented to Parliament an important "White Paper" (Cmd. 6527) on Employment Policy. This envisages a considerable development of statistical activity dependent upon effective cooperation between the Government and industrial representatives. It is stated that "the Government appeal with confidence to industry to join with them in a task which is essentially cooperative; for only industry can provide the statistical information required, and only a central authority can clarify and analyze information drawn from the country as a whole." The Report states that the following are the principal classes of statistics (in addition to those available before the war) which must be obtained for the efficient operation of an employment policy:

(a) Statistics of employment and unemployment, including quarterly or monthly statements of present and prospective employment in the main industries and areas in the country, based on returns from employers.

(b) Regular information relating to savings, projected capital expenditure by public authorities, and, as far as possible, by private industry.

(c) An Annual Census of Production showing the structure of the main groups of industries in the preceding year, including, *inter alia,* details of the quantity and value of output, stocks, and work in progress.

(d) Monthly figures of production, consumption, and stocks and, if possible, figures of orders on hand, based on sample returns obtained periodically throughout the year from large firms, trade associations, and public institutions.

(e) Annual and quarterly estimates of foreign capital movements and balance of foreign payments.

It is thus evident that there will be abundant scope in future for the activities of the Central Statistical Office, if it may be assumed that the "Central Authority" above mentioned should be the Central Statistical Office already in existence.

A Statistics of Trade Bill, promoted by the Board of Trade, was read a second time in the House of Commons on January 21, 1947, and received the Royal assent on July 31, 1947. It provides for the collection of statistical information from industrial undertakings, for an Annual Census of Production from 1948 and for a Census of Distribution and other Services in any calendar year as may be prescribed by Order of the Board of Trade. The Central Statistical Office is not mentioned specifically in the Act; but doubtless the Office will take due cognizance of the result obtained under the powers it confers.

A paper on "Statistics in Peace and War" was read before the Manchester Statistical Society on June 14, 1944, by Mr. H. Campion, who has been Director of the Central Statistical Office since May 1941. In this paper, Mr. Campion reviewed the statistics up to the outbreak of the late war; described the development of statistics since the outbreak of the war, and referred to statistics after the war. Respecting the recommendations of the Royal Statistical Society given above, he writes:

> "These proposals are still under consideration by the Government, but already the Chancellor of the Exchequer has stated in the House of Commons that some form of statistical organization will be kept in existence. Further, in the White Paper on Employment Policy, the Govermment announces its intention to establish a central staff to measure and analyze economic trends to help in the direction of a full employment policy. It now seems clear that the Government intend to take a lead in developing statistical work after the war and that the statistical service of the Government will be considerably strengthened."

The *Economist* of September 5, 1944, had an article on the Coordination of Statistics, from which the following is an extract:

> "The machinery of coordination is also already in being embryonically. In 1941 the Central Statistical Office was formed to reduce the mass of departmental data to terms that could be of service to the War Cabinet.... There must be a real coordination; and it is a well-tested principle that coordination works most smoothly and most efficiently if those whose work is to be coordinated are brought together to do the coordinating. Efficient coordination could most easily be achieved by continuous cooperation between the Central Statistical Office and a Committee chosen from the heads of the departmental statistical branches."

Reference has already been made to the various annual Statistical Abstracts of the Board of Trade and to their suspension during war; but it has been recently announced that the *Statistical Abstract for the United Kingdom* is now to be resumed, and is to be published in 1947, for the first time, by the Central Statistical Office instead of by the Board of Trade. The one last issued gave for the fifteen years, 1924 to 1938, the statistics available over a wide field including: climate; population; social, civil and industrial conditions; defense; finance; agriculture; fisheries; mining; metals; transportation; shipping; trade and commerce, etc.

CONCLUSION

The fact that a Central Statistical Office, so long advocated, is now in existence appears to open up important possiblities of future development. From what has already been stated, it is clear that there is abundant scope for the future activities of the existing Central Statistical Office, which the Government has apparently decided definitely to maintain. In conclusion I would submit that one direction which its activities might well take is the publication of a *Statistical Year Book*, as indeed was suggested by Geoffrey Drage in his paper of December 19, 1916, already mentioned. Such a publication should have adequate explanatory letterpress of tables over a series of years to illustrate progress in all departments of the public life. Special articles descriptive of the British systems of central and local Government, education, National Health Service, etc., would be of value. Useful publications of this kind have long been issued by the Dominions, notably by the Bureau of Statistics of the Dominion of Canada and by the Census and Statistics Office of the Commonwealth of Australia.

Résumé

L'auteur donne d'abord un aperçu des efforts faits, pendant de nombreuses années, par les statisticiens britanniques en vue de l'amélioration des statistiques officielles par la création d'un Office Central de Statistique.

Sous l'influence des besoins urgents causés par les circonstances de guerre un tel office a été créé en 1941.

L'auteur fait un exposé de la composition et du fonctionnement de l'Office Central et donne des détails sur ses publications. Il trace, dans ses grandes lignes, un plan de développement futur.

Discussion

Mr. Leak remarked that Mr. Godfrey's paper should be regarded as essentially a historical document. Mr. Idenburg commented on the significance of the British experience in developing centralization and coordination of statistical work.

UNE FAMILLE DES CLASSES MOYENNES D'UNE GUERRE MONDIALE A L'AUTRE, 1914-1947

par Ugo Giusti
(*Italie*)

Dans nos recherches sur les conditions des différentes classes sociales on a bien difficilement l'occasion de disposer d'une histoire complète et sincère d'une famille des classes cultivées, pendant une période assez longue. C'est que les hommes ouvrent plus facilement en prose et en vers, aux regards du monde, les endroits les plus cachés de leur coeur et les passions qui les agitent, que les chiffres les plus innocents de leur budget domestique.

D'ailleurs, si des documents de ce genre sont toujours une source précieuse d'information pour l'économiste et le sociologue, le statisticien, au contraire, ne peut en profiter que s'il connaît parfaitement les changements de tout ordre survenus dans la famille dont il a les données.

J'ai donc profité bien volontiers de la confiance d'un de mes amis qui avait eu la patience de suivre régulièrement et avec la plus grande précision, pendant plus de trente ans, sa comptabilité familière, pour en tirer un tableau ou se reflètent les grands événements qui ont changé deux fois de suite d'une façon radicale les conditions économiques d'une grande partie de la population italienne.

De cette étude monographique encore inédite je ne prendrai ici que quelques détails des plus significatifs dans l'espoir qu'ils soient suffisants pour donner une idée de ce qu'est devenue à l'heure actuelle, dans mon pays, une partie importante des classes moyennes.

(A) PENDANT LA PREMIERE GUERRE MONDIALE

La famille qui, de 1914 à 1921, résidait à Florence, se composait de cinq membres: le père de famille, un fonctionnaire administratif de 41 ans, un garçon de 14 et trois femmes respectivement de 74, 46, et 18 ans. Son budget se présentait au commencement et à la fin de la guerre de la manière suivante:

Revenu des Années

1914			1918		
Traitement	**Lires**	**5,000**	**Traitement**	**Lires**	**6,000**
Intérêts		**3,000**	**Indemnité de vie chère**		**1,100**
			Intérets		**1,500**
			Activité personnelle de la mère de famille		**1,500**
	Lires	**8,000**		**Lires**	**10,100**

C'est seulement à la moitié de février 1918 qu'on a pu bénéficier d'une indemnité de vie chère de 18% sur le traitement tandis que le chiffre-indice du coût de la vie (100 à la fin de décembre 1913) était déjà monté à 108 en décembre 1914, à 138, 150, et 224 respectivement à la même date de 1915, 1916, et 1917 pour continuer sa course jusqu'à 501 en 1921, terme extrême, pour ce moment, de sa montée. Pendant ces années, la famille a dû liquider une partie de ses épargnes et la mère a donné des leçons. Mais ni ces nouvelles ressources, ni les réductions radicales des dépenses que nous verrons bientôt, n'ont suffi à maintenir l'équilibre du budget:

1914		1918	
Revenus Dépenses	Lires 8,000 6,492	Revenus Dépenses	Lires 10,100 10,597
Solde actif (+) ou passif (—)	+ Lires 1,508		— Lires 497

La distribution des dépenses dans les deux années était assez différente:

	1914 %	1918 %
Alimentation	40·0	56·0
Eclairage et chauffage	2·8	5·2
Logement (loyer)	11·6	7·1
Produits hygiéniques et blanchissage	1·8	2·6
Vêtement et ameublement	6·1	7·5
Médecin et médicaments	1·4	1·9
Service domestique	4·6	4·5
Villégiature, voyages, etc,	9·3	—
Diverses	5·9	2·3
Impôts et taxes	16·5	12·9
	100·0	100·0

Les dépenses de villégiature, de voyages, etc., ont disparu en 1918; le coût du loyer, maintenu par le moratorium bloqué pendant la guerre, a perdu dans le pourcentage plus de quatre points; l'alimentation, au contraire, en a gagné seize malgré la grande diminution en poids et en qualité des produits consommés:

Consommation quotidienne par unité*

		1914		1918
Pain et pâtes alimentaires	Kg	0,560	Kg	0,381
Légumes (verts et secs) et fruits		0,394		0,490
Viande de toute espèce		0,148		0,077
Lait, beurre, fromage, oeufs		0,676		0,575
Sucre		0,066		0,023
Huile d'olive		0,030		0,015
Vin		0,166		0,093
	Kg	2,040	Kg	1,654

* Le coefficient pour l'homme adulte étant fixé à 1, on a appliqué celui de 0.75 pour les femmes et le garcon. Il faut pourtant observer que si, non sans quelques réserves, on peut accepter des échelles de ce genre quand il s'agit de comparer de grandes masses de consommateurs, elles ne sont pas applicables à des groupes d'individus plus limités et bien connus dans leurs habitudes et dans les caractéristiques de leurs consommations, et, moins encore dans la monographie d'une famille, sans tenir compte des conditions réelles physiques et sanitaires des membres de la famille étudiée.

Traduite en valeurs nutritives et calorifiques, cette consommation donnait les résultats suivants:

		1914		1918
Protéines	gr	90	gr	75
Corps gras		155		76
Carbohydrates		478		335
Calories (vin exclus)		3,761		2,391

(B) ENTRE LES DEUX GUERRES

Après les difficultés du passage de la guerre à la paix; l'année agricole de 1923 exceptionnellement favorable permettait à l'Italie de reprendre haleine et de maintenir les progrès déjà acquis, contrairement à ce qui arriva après la fin de la guerre récente et l'année agricole de 1945 exceptionnellement mauvaise.

Les effets de la déflation monétaire, de la crise mondiale et, à l'approche de la seconde guerre, ceux de l'autarchie poursuivie par le gouvernement fasciste, se réflètent dans le mouvement des nombres-indices des prix de gros et, moins vivement, dans celui du coût de la vie:

Nombres-indices (1922 = 100)

Année	Prix de gros	Cout de la vie	Année	Prix de gros	Cout de la vie
1923	100.9	99.4	1932	64.9	89.4
1924	100.4	102.9	1933	59.1	85.5
1925	112.5	115.6	1934	57.8	81.2
1926	114.8	124.7	1935	63.6	82.4
1927	96.5	114.0	1936	71.2	88.8
1928	93.2	105.7	1937	83.0	97.6
1929	88.9	107.6	1938	88.8	105.4
1930	79.6	103.9	1939	92.6	110.1

Les dépenses alimentaires de la famille suivent de près les nombres-indices nationaux; si les dépenses de 1928 sont faites égales à 100, on a pour les années de 1932 à 1937:

1932 — 81.7	1934 — 74.6	1936 — 83.4
1933 — 77.3	1935 — 76.2	1937 — 90.5

Les économies familiales qui ne produisent pas des biens économiques, ont ressenti, en sens inverse des économies de production, la crise des prix dans les marchés internationaux; notre famille s'est trouvée soulagée, pendant ces années de bas prix, des pertes subies pendant la guerre et l'après-guerre immédiate, mais ce soulagement a été bien peu de chose en comparaison aux désastres que devaient apporter la nouvelle guerre et ses effets directs et indirects.

A titre d'exemple, nous donnons le budget de 1934. La famille, qui, dès 1922, résidait à Rome, était restée composée de quatre membres: deux hommes (de 61 et de 33 ans en 1934) et de deux femmes (de 69 et de 24 ans). Ses revenus se composaient à cette date d'une pension de retraite de 12,000 lires, d'un traitement de 24,000, d'un loyer figuratif de 10,000 lires pour un appartement de propriété; au total, 46,000 lire. Le total des dépenses étant de 42,986 lires, il y avait un excédent actif de 3,014 lires.

La classification des dépenses était la suivante:

	%		%
Alimentation	20.6	Médecin et médicaments	3.4
Eclairage et chauffage	3.8	Service domestique	4.8
Logement (loyer figuratif)	25.3	Villégiature voyages, etc.	16.4
Vetement et ameublement	9.5	Diverses	5.9
Produits hygiéniques et blanchissage	1.1	Impôts et taxes	9.8

La consommation quotidienne par unité est indiquée ci dessous en poids de produits consommés, en valeurs nutritives, et en nombre de calories:

Pain et pâtes alimentaires	Kg	0.330	Protéines	gr	107.2
Légumes verts et secs, fruits		0.572	Corps gras	gr	123.6
Viandes de toute espèce		0.107	Carbohydrates	gr	325.8
Lait, beurre, fromage oeufs		0.470			
Sucre		0.019			
Café		0.010	Calories		2,921
Huile d'olèvi		0.029			
Vin		0.267			
	Kg	1.804			

En comparaison avec l'année 1914, on peut remarquer en 1934 une diminution très forte dans le pourcentage des dépenses pour l'alimentation (20.6 au lieu de 40) due en partie au changement réel dans la quantité et dans la qualité des produits consommés (236 gr. de moins par jour; légumes au lieu de laitages, de sucre, etc.) et, indirectement, au plus haut prix des loyers, maintenus par le moratorium à Florence en 1918 et libres à Rome en 1934, et aux plus fortes dépenses pour villégiature, voyages, etc.

(c) PENDANT LA SECONDE GUERRE MONDIALE

Si des décrets du gouvernement ou, en général, des règlements gouvernementaux pouvaient régler la vie des sociétés humaines, nous pourrions dire que le gouvernement italien avait bien protégé ses administrés dès le commencement de la guerre, des dangers du renchérissement du coût de la vie. "Aucune augmentation," disait un décret de juin 1940, "ne sera consentie sur les prix en vigueur pour les marchandises de toute espèce, les services, etc,"; quelques centaines de décrets qui ont suivi admettaient tout au plus des exceptions et garantissaient pour le reste.

On est tenté de rechercher ce que nous indique à cet égard le chiffre-indice officiel du coût de la vie à Rome (1928 = 100):

1939 — 107·28
1940 — 125·37
1941 — 144·14
1942 — 169·82

Après 1942 la voix officielle a gardé le silence; c'était le mieux qu'elle pouvait faire malgré la précision jusqu'aux centimes dont elle faisait parade; la comptabilité de notre ami nous dit en effet qu'en juin 1942 il a dépensé 1101 lires pour 75 kilos de produits alimentaires au lieu de 448.90 pour 97 kilos au même mois de 1939; tout simplement parce qu'un kilo de viande de boeuf, indiqué sur la mercuriale officielle L. 18.78, lui a coûté 81.25; un kilo de beurre 100 lire au lieu de 27 et ainsi de suite pour les autres produits.

Nous n'avons pas le temps de faire des commentaires sur l'action des pouvoirs publics dans cette matière, mais nous en verrons les effets négatifs à travers les chiffres du budget dont nous disposons.

Ce sont d'abord les revenus qui tâchent en vain de tenir tête aux dépenses; au lieu du traitement de 24,000 lires, on voit en 1940 une contribution de 15,000 du fils de famille, 10,000 lires d'intérêts, et 5,000, produit de l'activité du chef de famille; tout ça ensemble fait 52,000 lires.

Mais les prix, malgré gouvernemental cité plus haut, montaient déjà sensiblement et les dépenses pour l'alimentation passaient de 9,862 en 1934 à 12,228 lires en 1940; des économies sur les dépenses pour la villégiature et sur quelques autres chapitres ont permis de rétablir l'équilibre en réduisant le total des dépenses à 45,545 lires, etlaissant un solde actif de 6,455 lires.

Après quatre ans, nous trouvons une situation bien différente; les revenus ont monté de 52,000 à 89,000 à la suite d'une allocation de vie chère et d'un redoublement d'activité personnelle, mais les dépenses ont monté plus vite de 45,565 à 193,931 lires, en laissant un déficit de 104,931 lires.

Le simple pourcentage des différents chapitres des dépenses dans les deux années nous dira éloquemment ce qui s'est passé dans l'intervalle:

	1940	1944–45
Alimentation	26.9%	71.2%
Eclairage et chauffage	5.2	6.5
Logement (loyer figuratif)	24.7	6.4
Vêtements et ameublement	11.2	2.0
Produits hygiéniques et blanchissage	1.9	3.0
Médecin et médicaments	2.1	1.7
Service domestiques	5.3	3.6
Villégiature, voyages, etc.	10.4	—
Diverses	7.3	3.4
Impôts et taxes	5.0	2.2

Les dépenses alimentaires qui correspondaient en 1940 à un quart environ de la dépense totale (pourcentage qu'on peut considérer normal pour une famille des classes moyennes dans des circonstances également normales), comprenait en 1944-45 presque les trois—quarts du budget passif malgré la forte augmentation absolue dans les autres chapitres, exception faite pour les dépenses de villégiature et des voyages qui ont été supprimées. C'est le pourcentage qui caractérise ordinairement les budgets des familles les plus misérables.

L'annexe nous présente l'histoire alimentaire de la famille dans les années de la guerre récente et lointaine; les 2,912 calories quotidiennes par unité consommatrice au mois de mars 1939 descendent à 1,633 en avril 1944 (actions de guerre près de Rome) pour remonter à presque 3,000 en décembre de la même année, et, après une nouvelle baisse, à la même quantité au commencement de 1947; mais c'est l'origine de ces calories aux différentes époques qui est bien différente: équilibre de principes nutritifs en 1939; manque presque absolu de produits alimentaires en avril 1944; forte consommation de céreales accompagnée par une absence presque totale de viande et de corps gras à la fin de la même année. La situation a fort peu changé à cet égard dans les premiers mois de 1947; la carte de rationnement n'y a joué qu'un bien faible rôle; si, en 1942, elle pouvait fournir 40% en poids de la consommation et 13% de la dépense relative, elle n'a donné depuis 1944, tout en comprenant un supplément d'infirmité, que de 13 à 20% en poids et de 2 à 4% des dépenses.

Malgré les réductions remarquables, surtout en qualité, des produits consommés et les préoccupations d'ordre économique qui les ont accompagnées, il est intéressant de constater que la santé de la famille et même celle de la vieille dame infirme, n'en ont pas souffert. Tous ont plus ou moins diminué de poids mais seulement le père de famille d'une manière évidente (21 kilos sur 80) avec quelques manifestations de faiblesse dans les derniers temps; la forte consommation de pain et de pâtes alimentaires achetés au marché libre lui ont fait assez rapidement reprendre son ancien poids.

Toutes ces réductions n'ont pourtant pas empêché, comme nous avons vu, l'énorme déficit de 100,000 lires qu'on a comblé en aliénant des épargnes es une quantité d'objets personnels et de meubles.

L'année 1946 qui aurait dû apporter, avec la paix, un peu plus de tranquillité et de soulagement, nous a donné une cruelle déception; les prix des denrées ont monté presque sans interruption; le budget mensuel de dépenses qui était de 16,161 lire pour quatre personnes en 1944-45 monta à 30,000 pour trois au commencement de 1947. Malgré l'augmentation en valeurs absolues de toutes les dépenses, l'alimentation absorbe toujours les trois-quarts du budget passif et exige au mois de mars 1947 plus de 300 lires quotidiennes par unité consommatrice.

La reconquête de mille calories entre 1944 et 1947 n'a pas fait oublier a différence radicale entre les repas d'aujourd'hui et ceux d'il y a quatre ou cinq années seulement; plus de beurre, de vin, de café; presque plus de viande; une cuisine égale, monotone, comme celle d'un établissement d'assistance publique a remplacé les jouissances, pourtant modestes, de la bonne chère familiale de jadis.

D'ailleurs le budget, qui n'est qu'un compte de caisse, ne nous dit rien du riche patrimoine de la famille qui va disparaître; ce sont des meubles qu'on ne peut plus entretenir, des vêtements, du linge, des souliers qu'on ne peut plus renouveler; la maison même où la famille habite lui est devenue un poids trop lourd pour sa condition actuelle et peut-être devra-t-elle l'abandonner; on tremble à la pensée d'une maladie; l'avenir se présente bien noir.

Il n'est ni difficile ni arbitraire de remonter de ce fait personnel (surtout si on peut s'en rendre compte, non dans ces notes rapides, mais dans la riche documentation de la monographie) au fait collectif qui se manifeste sous nos yeux. C'est la déchéance économique, et partant sociale, des classes moyennes, de toute une classe qui voit le travail d'une vie entière anéanti, qui, sans force de lutter contre les adversités, tombe jour par jour dans une misère d'autant plus douloureuse, cachée comme elle l'est par les restes d'un bien-être récent, par les habitudes d'une société cultivée, par la répugnance de ces nouveaux pauvres aux protestations bruyantes.

Annex

Consommation quotidienne, calories, dépenses par unité consommatrice, coût de 1000 calories

Date		Consommation (grammes)	Calories	Dépense (Lires)	Coût de 1000 calories (Lires)
Juillet	1914	2,040	3,876	1.68	0.43
Août	1918	1,654	2,456	3.51	1.43
Juillet	1928	1,698	2,911	9.26	3.18
Sept.	1932	1,781	2,966	7.63	2.59
Sept.	1935	1,804	2,921	7.10	2.43
Sept.	1937	1,774	2,949	8.83	3.00
Mars	1939	1,835	2,912	8.57	2.95
Juin	1942*	1,407	2,121	21.20	10.00
Juin	1943	1,597	2,443	18.80	7.71
Octobre	1943	1,327	2,611	31.20	12.00
Novembre	1943	1,388	2,344	44.00	18.80
Janvier	1944	1,125	2,304	44.30	19.30
Avril	1944	956	1;633	108.00	66.30
Mai	1944	1,014	1,827	158.30	86.50
Juillet	1944	1,700	2,366	96.50	40.70
Décembre	1944	1,449	2,993	104.00	34.60
Mars	1945	1,339	2,624	103.00	39.30
Janvier	1947	1,800	3,005	244.00	81.30

*A partir de 1942 et pendant des mois, certains produits alimentaires manquaient totalement sur le marché; ces changements d'alimentation apportaient de fortes différences dans le rendement en calories. Voir, par exemple, l'effet de la forte consommation de légumes et de fruits secs en décembre 1944.

Résumé

We seldom have available the economic history of a family for a period of thirty years—that is through the two world wars—as is the case for the Italian family of the middle classes of which this paper gives a brief account. Particularly interesting with regard to this subject is, for instance, a comparison between the family budget of 1940, at the beginning of the war when it was still possible to save 10% of income, and

the budget of 1944-45 when the family was obliged to sell 100,000 lires of funds and commodities to meet expenditures. The distribution of funds was of course completely different at those two dates; only 26.9% for food in 1940 (a normally characteristic percentage for the middle-class standard of living) and almost three-quarters of the whole expenditure (71.2%) four years later, that is, the usual percentage for families, in the most wretched economic condition. The expenditures for travel theater, etc., had disappeared from the budget in the meantime. The great want of food products in the midst of the war (for Rome, April 1944) caused a reduction of the normal number of calories (from 2900 to 1633 per person per day); already in December 1944 the former number was recovered but now it corresponded to a wholly different sort of diet: no more butter, wine, coffee, seldom any meat; a tedious diet based on bread, maccaroni, eggs, and vegetables had taken the place of the pleasant variety of meals available formerly. Pleasure was greatly reduced but the health of the family did not suffer excessively from this change. It is easy to pass in the consideration of these facts from an individual to a collective conclusion: a large part of our middle classes looks, without possibility of defense, at the ruin of all the results of its lifework and suffers particularly under its misfortune just because that misery is often concealed under the remainder of recent prosperity.

OSSERVAZIONI SUL METODO DI RILEVAZIONE SUL VOLUME DELLA CIRCOLAZIONE SOCIALE

per Livio Livi
Università di Firenze (Italia)

Si parla molto dei problemi relativi alla circolazione sociale, intendendo con questo termine quella parte del flusso di rinnuovo del ceto più elevato della società costituito da persona provenienti dal ceto dei lavoratori manuali; ma in realtà ben poco si conosce del volume di questo movimento e delle vie che esso segue. Parzialissime e scarse sono anche le ricerche sul grado di selezione cui sono soggette le persone che provengono ai posti più elevati dagli strati più bassi, rispetto a quelle che hanno origine da famiglie già stabilizzate nel ceto medio o alto.[1]

Eppure, dopo i primi ben noti studi del Fahlbeck,[2] l'argomento é stato ripreso più volte ma sempre in modo parziale e con riferimento a quella parte della circolazione che si sviluppa per il tramite della scuola superiore.[3]

---

[1] Abbiamo solo conoscenza dei seguenti scritti: T. M. Livesay, "The Relation of Test Intelligence of High School Seniors in Hawaii to the Occupation of Their Fathers," *Journal of Applied Psychology*, 25, 1941. F. Zeuthan, "Professione o posizione sociale sono ereditarie?" *Socialisten*, 1933.

[2] P. Fahlbeck, "Les classes sociales," *Bull. de l'Inst. Int. de Stat.*, XIX, 3me livr.

[3] Tra i più recenti citiamo: H. B. Goetsch, *Parental Income and College Opportunities*, New York, 1940. Olot, "Conditions économiques des étudiants," Helsinki, Bureau Centr. de Stat., 1939, edizione bilingue finnico e svedese con riassunto francese. Vincenzo Castrilli, "Origine sociale des étudiants," *Bull. de l'Inst. Int. de Stat.*, 1940. Académie de Paris, *Enquetes sur le Baccalauréat*, Paris, 1936. G. Deyerler, "Soziale Schichtung und schulische Leistung der volksschulpflictigen Yugend Münchens," "*Bayer. Stat. Landesamte*, 1936. G. Just, "Schulauslese und Lebensleistung" *Arch. f. Bevolkerungwissenschaft*, XII, 1935. Office de Statistique de la Rep. Tshecoslov., *Conditions sociales des étudiants des écoles supérieures de Brno*, 1935. R. Lotze, *Untersuchungen uber die Beziehungen zwischen Schulleistungen, sozialer Schichtung und Familiegrosse*," Stuttgart, 1934. A. O. Molnar, *Le condizioni sociali ed economiche degli studenti di scuole superiori a Budapest*, Bureau de Stat. de la Ville de Budapest, 1934 (in ungherese). J. Wisniewski e E.Violrose, *Risultati del l'inchiesta sulla situazione economica degli studenti delle scuole centrali di Commercio di Varsavia*," Varsavia, 1938 (in polacco). M. Miller, *The Generations, A Study of the Cycle of Parents and Children*, London, 1938. P. Fortunati, "Il ricambio sociale secondo le indagini sulla struttura demografica della studentesca universitaria di Padova," *Annali dell'Universita di Ferrara*, 1937. S. P. Ratz Kotsicas, "L'orientation professionelle des étudiants en Grèce, par rapport à la profession du père," *Bull. de l'Inst. Int. de Stat.*, 29, 1937. D. Krejci, "Expériences faites en Tchécoslovaquie au cours des relevés statistiques concernant la condition sociale des étudiants," *Bull. de l'Inst. Int. de Stat.* XXIX, 1937.

Quegli studi poi che non si pongono questo confine si sono dovuti restringere per la mancanza delle fonti statistiche allo spoglio di nomi registrati in dizionari biografici o di persone ospitate in istituti speciali,[4] oppure si sono basate sopra elementi indiretti tratti dai censimenti domografici.[5]

Noi stessi, come vedremo, per avere una idea della importanza quantitativa della circolazione sociale nella categoria degli imprenditori industriali e commerciali, abbiamo dovuto far ricorso ad un annuario della società per azioni.

E' doloroso per gli statistici essere costretti a coltivare un campo così vasto e fertile con attrezzi così meschini.

Si aggiunge poi l'inconveniente che le cifre così raccolte si riferiscono per solito a concetti molto confusi, e sono perciò mal comparabili. Si parla di ceto alto o direttivo, e di ceto dei lavoratori manuali, ma é ovvio che a questi concetti possono darsi definizioni diverse e tutte molto imprecise per la stessa complessità dell'oggetto. Esporremo nel paragrafo successivo un metodo che non renderebbe necessarie queste equivoche definizioni, ma per ora non siamo potuti andare al di là di una sempliceproposta.

Inoltre il volume di questa circolazione, anziché sui fenomeni di flusso, può essere studiato, come per lo più avviene, sui fenomeni di stato, cioè in base alle frequenze così come appaiono in un dato istante nella collettività considerata. Ma i risultati sono diversi secondochè, della circolazione socilae si considera l'aspetto dinamico e l'aspetto statico.

Vi è poi da precisare il termine "a quo" del passaggio. Si può passare dal basso ceto ai posti direttivi nel corso stesso della vita. Potremo chiamare questi passaggi di 1° grado. Il Prof. Luzzatto-Fegiz, pur non potendone dare una misura, ha dimostrato che questa circolazione è notevole.[6]

Seguendo la linea maschile vi sarebbero poi passaggi di 2° grado da padre a figlio; di 3° grado da nonno a nipote, e così via; ma non converrebbe certo andare più oltre.

Dovrebbesi anche stabilire se e come tener conto della discendenza femminile; sarebbe anzi interessantissimo comparare la circolazione sociale maschile, cioè calcolata per i maschi seguendo la linea paterna, e quella femminile, cioè calcolata per le donne seguendo la linea materna. Le differenze potrebbero essere notevoli.

---

[4] "Sozialer Auf—und Abstieg im duetschen Volke," *Bayer. Stat. Landesamt,* 1930. A. Odin, "Genèse des grands hommes," Paris.

[5] Taussing e Joslyn, *American Business Leaders,* New York, 1932. Pierpaolo Luzzatto-Fegiz, "The Occupational Evolution of a Generation," *Population,* Vol. II, 1935.

[6] Pierpaolo Luzzatto-Fegiz, *Op. cit.*

In sostanza l'ostacolo prinicipale di queste ricerche sta nella difficoltà di ordinare le persone attive di una data collettività secondo il grado di importanza che esse hanno nel gruppo stesso, date che questa importanza non può desumersi soltanto dal reddito goduto o dalla professione esercitata, ma dal complesso di queste e di altre circostanze. Ma supponendo superata questa difficoltà ed ottenuta una graduatoria anche non esattissima, tutte le imperfezioni ed equivoci derivanti dalle definizioni dei gruppi costituenti il ceto alto o il ceto basso. verrebbero ad essere superati; non vi sarebbe infatti necessità di ricorrervi.

Si supponga di voler considerare la circolazione sociale nel suo aspetto statico; graduate le persone attive componenti un dato gruppo sociale secondo la loro importanza, si potrà procedere ad una analoga graduazione dei padri di queste persone (o di altro ascendente)[7] ed allora potrebbe prestarsi assai bene per la rappresentazione e la misura della circolazione sociale un procedimento analogo a quello chec si applica per lo studio della concentrazione.

Se, a cominciare dall'alto della graduatoria, si delimitano gruppi percentuali sempre più grandi abbassandone il limite inferiore, si potrà per ciascuno di questi gruppi determinare la percentuale di coloro che provenivano da padri (o altri ascendenti) che, nella loro graduatoria erano compresi nello stesso gruppo percentuale.. Per brevità indicheremo questi col termine di "omosociali."

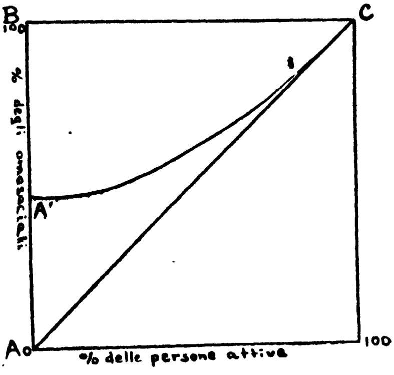


Grafico 1

Si riportino sull'asse delle ascisse di un diagramma cartesiano le percentuali dei singoli gruppi formati nel modo che si é detto, e sull'asse delle ordinate si riportino le percentuali degli omosociali, come é indicato nel grafico 1.

---

[7] Più esattamente, verrebbero graduati in simile modo i *casi di paternita* cioè gli stessi appartenenti al gruppo, ma non più in riguardo della loro posizione sociale, ma bensi della posizione sociale del padre (o di altro ascendente).

Con prima approssimazione, cioè trascurando l'effetto di altre cause di perturbazione,[8] si potrebbe dire che, se la circolazione sociale avvenisse come in un gioco di sorte, cioè senza che operino fattori ambientali, ereditari o selettivi, a ciascuna frazione delle persone attive considerate corrisponderebbe una egual frazione di "omosociali." Le percentuali di questi ultimi verrebbero a disporsi sulla diagonale BC.

Ma in realtà queste percentuali si pongono sopra una linea curva A'C più sollevata, e di tanto più sollevata, quanto minore è il processo di osmosi sociale. La superficie racchiusa tra la diagonale BC e la curva A'C darebbe una misura della occlusione sociale.

Nel caso limite di totale occlusione, tale superficie risulterebbe eguale all'intero triangolo BAC, sicché il rapporta delle due superfici $\frac{A'AC}{BAC}$ potrebbe darne una misura; mentre il rapporto complementare $\frac{A'BC}{BAC}$ indicherebbe il grado di osmosi sociale.

Naturalmente la diagonale BC sarebbe ben lungi dal rappresentare una situazione di "ottimo." Anche amesso una ambiente avente una piena libertà di circolazione, vi sarebbero pur sempre circostanze naturali ineliminabili come la selezione bioereditaria dei figli dei più altolocati, ed altre circostanze che sarebbe dannoso eliminare (come la migliore educazione e preparazione darivanti da un ambiente familiare più elevato). Tutto ciò comporterebbe, pure amessa l'eliminazione di tutti gli attriti di carattere patologico, una forte deviasione del segmento BC verso l'alto.

Ma non v'è dubbio che una gran parte della superficie OA'C deriva da privilegi eliminabili e dannosi.

Per tutte le difficoltà di cui abbiamo parlato i saggi dovuti a privati studiosi che hanno trattato questo argomento, sono eseguiti in modo estremamente grossolano. Per solito si considera un gruppo di persone appartenenti all'alto ceto, ma anzichè determinare qual percentuale di questo gruppo proviene oppur non da ascendenti che avevano la stessa possizione relativa, si determina la percentuale di coloro che provengono da un gruppo assai più ampio, per solito contrapposto a quello dei lavoratori manuali.

---

[8] Con questo metodo non si prende in considerazione quella parte del flusso dovuta alla sostituzione di quegli elementi che non ebbero prole maschile attiva. Questa esclusione conduce ad una attenuazione nella misura della circolazione tanto più forte quanto maggiore è la frequenza delle persone che rimangono senza discendenti attivi, appartenenti all' alto ceto, rispetto alla analoga frequenza propria, dei ceti più bassi. Altra causa di deformazione consisterebbe nel considerare la condizione sociale degli appartenenti al gruppo in esame all'eta che essi hanno al tempo del rilievo, e considerare invece la posizione sociale dei padri ad una età o ad un periodo della vita eguale per tutti. Queste cause di perturbazione possono essere evitate o almeno in parte corrette.

Riferiamo questo tipo di indagini al precedente schema (vedi grafico 2). Il segmento Aβ indica l'ampiezza percentuale del gruppo superiore considerato; il segmento Aα, l'ampiezza del ceto dei lavoratori non manuali. Poichè le singole ordinate indicano la percentuale di coloro che provengono non già da padri aventi la loro stessa posizione, ma da padri aventi una posizione anche assai più modesta, la percentuale degli "omosociali" si solleverà (dal punto A″ al punto A‴), e tanto più si solleverà, quanto più piccolo é il segmento Aβ , e quanto più grande è il segmento Aα.

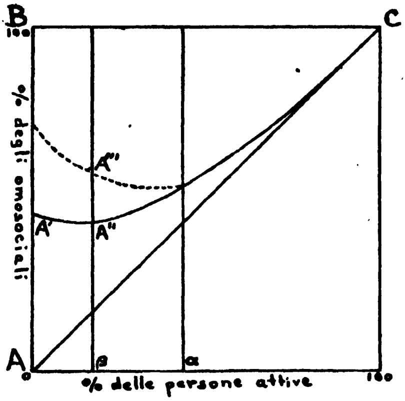


Grafico 2

Per il fatto che tale percentuale è funzione sia dell'ampiezza che si è data al gruppo superiore o direttivo, sia dell'ampiezza del gruppo dei lavoratori non manuali, deriva la equivocità e la incomparabilità di questi risultati. Non di meno anche noi, nel breve saggio che sarà esposto in seguito, abbiamo dovuto accostarci al comune ed imperfetto schema.

L'ampiezza del segmento Bα può dipendere, non solo dalla definizione che si dette al "ceto medio," ma anche da cause oggettive. E' ovvio che anche se fosse attuabile una perfetta eguaglianza di definizioni, con questi vecchi schemi di indagine si otterrebbero risultati ben diversi a seconda che si tratti di paesi in cui il ceto manuale é, relativamente, più o meno frequente.

Non crediamo però che le difficoltà che si frappongono alla esecuzione di rilievi del tipo che abbiamo sintetizzato nel grafico 1, siano insuperabili.

Non è certamente questo un compito che possa essere assunto dalla Statistica ufficiale, sarebbe pur molto gravoso per privati studiosi. Ma l'esistenza di Enti che operano col metodo dei sondaggi renderebbe possibile l'esecuzione di indagini del tipo che suggeriamo. Sebbene la graduatoria che si dovrebbe formare (sia per le persone costituenti il campione, sia per gli ascendenti di queste) sarebbe pur sempre affetta da elementi soggettivi ed arbitrari, queste indagini darebbero risultati assai più completi e soddisfacenti in confronto di quelli di cui disponiamo al presente.

Queste ricerche hanno preso in oggi un interesse eccezionale anche per i riflessi politici del problema. Come nel campo della produzione si hanno sistemi di economia pianificata, e di economia di concorrenza, così nell'impiego delle capacità individuali si attuano o si vorrebbero attuare sistemi di circolazione pianificata o di circolazione libera.

Nel primo sistema il passaggio ai posti più alti dovrebbe avvenire sotto il controllo dello Stato; nel secondo sistema esso si verifica in modo spontaneo. Nel primo la selezione è fatta dallo Stato stesso con una scelta che si inizia dall'età scolastica; nel secondo la selezione è affidata alla libera concorrenza. Ma è ovvio che tanto nel caso della circolazione libera quanto nel caso della circolazione pianificata, si verificano o si possono verificare circostanze che ne attenuano il valore sellettivo.

E' ben noto che nei paesi a circolazione libera queste circostanze sussistono e sono di carattere ambientale ed economico; ma nessuno si e ancora avvicinato a darne una misura. Nei paesi a base comunista sarebbe poi da indagare se i criteri selettivi che si adottano sono i migliori, o se, nel corso del tempo non vengano a riformarsi dei previlegi particolari.

A tutti questi interrogativi dovrebbe dar risposta la Statistica con indagini approfondite. Nei paesi democratici queste indagini dovrebbero essere preliminarmente dirette a valutare il volume di questo flusso e a precisare le vie che gli "homines novi" seguirono nella loro ascesa. Cioè se furono formati principalmente dalla scuola o dalla pratica. Si dovrebbe poi indagare sull' esistenza e sul grado della selezione ereditaria e sugli effetti anti selettivi dell'ambiente.

L'idea che in uno Stato progredito tutti i posti di direzione e di responsabilità debbano essere occupati dai più capaci, indipendentemente dalla loro origine sociale, é accolta come principio indiscutibile ed è anche affermata come un diritto.

Ma la generalità giudica questo colossale problema in base al falso presupposto che la via maestra della elevasione sociale sia quella della scuola superiore. Sicché la soluzione starebbe tutta nell'offrire ai giovani dei ceti più bassi quelle stesse possibilità di accedere agli alti gradi dell' istruzione che sono offerte ai figli delle persone più altolocate.

Se si guarda soltanto a quello che accade nell'ambito universitario si può avere l'impressione che la circolazione sociale, nei paesi in cui il flusso di rinnuovo é rimesso prevalemtemente all'azione della libera concorrenza, incontri ostruzioni quasi proibitive. Infatti in Italia, come a un dipresso negli altri paesi occidentali d'Europa, la proporzione dei nuovi laureati aventi il padre di condizione operaia o assimilabile all'operaia, é circa del 12% del totale; e se si potesse poi calcolare la proporzione di coloro che, ottenuta la laurea, proseguono nell'ascesa per raggiungere posizioni di riguardo, si troverebbero certo ben più piccole percentuali.

Ma le cose stanno in modo ben diverso da come sono viste dalla generalità, forse domiata dal prestigio che circonda l'alta cultura. Oggetto principale di questa memoria é quello appunto di dimostrare, sebbene con piccolo saggio statistico, che la circolazione sociale non passa che in piccola parte dalla scuola universitaria, e che il suo volume, rispetto alla totalità del flusso di rinnuovo, è almeno triplo di quello che appare considerando soltanto l'origine familiare dei nuovi laureati.

Tutto questo ha un notevole interesse anche pratico poichè se per iniziativa dello Stato si vuole fare qualche cosa di utile occorre avere una idea meno imprecisa di quanto avviene in realtà e favorire quelle forze che già sono in atto spontaneamente.

Per questa indagine abbiamo preso le mosse da uno spoglio dei nomi delle persone di sesso maschile appartenenti ai consigli di amministrazione di Società per azioni che esplicavano attività industriali e commerciali[9] e che nel 1939 avevano un capitale sociale non inferiore ad 1,000,000 di lire.[10]

Abbiamo classificato questi nominativi secondo chè risultavano, oppur no, in possesso di una laurea o di altro titolo equipollente, distinguendo pure tra gli uni e gli altri quelli che risultavano iscritti nell' albo della "Nobiltà."

Non ci siamo sobbarcati al gravoso lavoro della eliminazione dei duplicati che sono frequentissimi.[11] Ma, come si vedra, tale omissione non impedirà di trarre dalle cifre raccolte conclusioni sicure, sebbane generiche.

---

[9] Più precisamente abbiamo considerato le industrie estrattive, metallurgiche, meccaniche, elettriche, tessili, dell'abbigliamento e del cuoio, del legname, degli agglomeranti, laterizi, ceramica, vetro, alimentari e della pesca, della carta e poligrafiche, commerciali (grandi magazzini, commercio all'ingrosso, importazione e esportazione).

[10] Abbiamo utilizzato per tale spoglio il volume: *Le Societa Italiane per Azioni—Notizie Statistiche,* Edito dalla Ass. Ital. fra le Soc. per Azioni; ed. XVI, Roma, 1940.

Abbiamo riparato almeno in gran parte alle rare omissioni dei titoli professionali e nobiliari con la consultazione di altri annuari e dizionari biografici. Per titoli equipollenti alla laurea abbiamo considerati quelli di Ufficiale superiore dell'Esercito, dell'Aviazione o della Marina; nonché la generica indicazione di "Professore"; criteri questi volutamente assai larghi che abbiamo adottato anche a compenso delle poche eventuali omissioni.

[11] Vedi P. Luzzatte-Fegiz, "Il consiglio di amministrazione e l'interdipendenza delle imprese," *Il giornale degli Economisti,* Marzo 1928.

Riportiamo anzitutto i risultati complessivi dello spoglio:

Laureati	4,518
Non laureati	4,959
TOTALE	9,477

I nomi di persone aventi titoli nobiliari si distribuivano invece nel mode seguente:

Laureati	296
Non laureati	187
TOTALE	483

Queste cifre benchè dimostrino la notevole frequenza di non laureati tra le persone che occupano elevati posti di comando nelle industrie e nei commerci, non dicono ancora nulla circa la entità della circolazione sociale, perchè ovviamente, e come la stessa distribuzione dei "nobili" dimostra, una gran parte dei non laureati deriva da famiglie già stabiliazzate nel ceto superiore. Nondimeno esse ci danno la possibilità di accostarci ad una valutazione di detta circolazione.

Abbiamo già detto tra i giovani laureati la proporzione di quelli provenienti dal ceto degli operai manuali è circa del 12%. Questo ci consente di affermare che sul totale dei laureati qui considerati (4,518) vi deve pur essere una frazione, non certo eguale, ma pur apprezzabile, di elementi sorti dal ceto operaio. Attenendoci ad una proporzione del 6% in luogo del 12%, non erreremo probabilment di molto. Avremo quindi: 4,518 × 0.06 = 271.

Poichè sappiamo che i laureati provenienti dal ceto nobiliare erano 296, si può dire che: 4,518 -- (271+296) = 3,951, indichi il numero dei laureati provenienti da famiglie già appartenenti al ceto medio. E se poi vogliamo ammettere che in questo ceto il rapporto dei non laureati rispetto ai laureati sia quello stesso che si rinviene nel ceto nobiliare (63·18%) si deve concludere che ai 3,951 laureati originari del ceto medio, devono corrispondere circa 3,951 × 0.6318 = 2,496 non laureati di eguale origine.

Abbiamo cosi tutti gli elementi per riempire la seguente tabella:

TABELLA 1

Italia, 1939: Distribuzione ipotetica dei Consiglieri dei Amministrazione di Società per Azioni aventi un milione o più di capitale sociale, secondo il ceto sociale di provenienza

Ceto di provenienza	Laureati	Non laureati	Totale
Nobiliare	296	187	483
Borghese	3,951	2,496	6,447
Operaio o assimilato	271	2,274	2,545
TOTALE	4,518	4,957	9,475

Da questa sfuocatissima impressione statistica risulterebbe che la precentuale degli "homines novi" sarebbe del 26.9%. Ma affinando il rilievo avremmo trovato una perentuale molto più alta poichè tutte le principali cause di perturbazione che infirmano questo rilievo relativamente al fine che si prefigge operano nell' abbassarla.

Infatti nel ceto borghese stabilizzato da almeno due generazioni ed occupante posti direttivi la frequenza dei non laureati rispetto ai laureati deve ritenersi più bassa di quella che si rinviene nel ceto nobiliare.

Si deve anche tener presente che la concentrazione dei seggi consiliari nelle stesse persone é verosimilmente di gran lunga minore per gli "homines novi" in confronto di tutti gli altri.

Inoltre tra i nobili, e specialmente tra i nobili non laureati, vi sono uomini il cui passaggio al ceto superiore è recentissimo.

Si deve infine tener presente che la predetta percentuale del 26.9 riguarda un fenomeno di *stato* e non di *movimento*, non riguarda cioè la composizione del flusso, ma la struttura del gruppo colta in un dato istante. Nel flusso di rinnuovo la percentuale degli "homines novi" e certamente maggiore di quello che si rinviene nella massa, e questo per un complesso di ragioni sociali e demografiche che ogni statistiche intuisce.

Naturalmente deve supporsi che tale percentuale sia più bassa nella parte più elevata di questo settore, cioè che decresca col crescere coll' importanza del capitale sociale. Ma dobbiamo anche riflettere che il nostro rilievo non coglie, nelle attività che abbiamo considerato, che una parte degli imprenditori, quella parte cioè che esplica tale attività nella forma meno personale delle Società per Azioni. In tutte le altre numerose aziende oltrepassanti il limite di importanza posto in questo studio ed aventi carattere personale o un'altra figura giuridica, la percentuale esprimente il volume della circolazione sociale sarebbe risultato, per ragioni intuitive, assai più forte.

Dobbiamo piuttosto riconoscere che in questa indagine non sono chiarite le caratteristiche qualitative di questa circolazione, non é precisato cioè se si tratta solo di figli di persone che esercitano professioni manuali o assimilabili ad esse, oppure se comprendono anche elementi la cui posizione superiore non è cosi recente.

Tutto però fa credere che la circolazione sociale costituita da passaggi di I° e di 2° grado sia in questo settore almeno tripla, nella sua importanza relativa, di quella che passa per la scuola universitaria.

Ma la nostra modesta ed imperfetta indagine dice qualcosa di più edi più esatto.

Avendo distribuito i consiglieri laureati e non laureati combinatamente secondo le varie attività industriali o commerciali, e secondo l'altezza del capitale sociale, siamo in grado di vedere le linee di maggior

TABELLA 2

Italia, 1939: Nominativi dei consiglieri di Amministrazione laureati e non laureati distribuiti secondo l'altozza del capitale sociale e le categorie industriali

Categorie industriali	da milioni 1 a 1.99			da milioni 2 a 4.99			di 5 milioni o più			In complesso		
	Laureati	Non Laureati	Totale	Laureati	Non Laureati	Totale	Laureati	Non Laureati	Totale	Laureati	Non Laureati	Totale
Estrattive	54	48	102	77	42	119	122	60	182	253	150	403
Metallurgiche	50	73	123	60	67	127	175	88	263	285	228	513
Meccaniche	206	210	416	296	224	520	454	239	693	956	673	629
Elettriche	70	47	117	54	30	84	575	168	743	699	245	944
Tessili	107	328	435	160	394	554	260	416	676	527	1,138	1,665
Abbigl. e Cuoio	25	66	91	15	62	77	24	28	52	64	156	220
Legname e affini	17	47	64	22	27	49	32	24	56	71	98	169
Agglom. Laterizi	103	170	273	111	125	236	159	97	256	373	392	765
Alimentari e Pesca	121	247	368	173	298	471	278	298	576	572	843	1,415
Carta e Poligrafici	97	95	192	53	39	92	129	74	203	279	208	487
Commercio	178	409	587	128	253	381	133	166	299	439	828	1,267
TOTALE	1,028	1,740	2,768	1,149	1,156	2,710	2,341	1,598	3,999	4,518	4,959	9,477
di cui "Nobili"	24	34	57	47	43	90	25	110	335	296	187	473
					Cifre proporzionali							
Estrattive	52.9	47.1	100	64.7	35.2	100	67.1	32.9	100	62.8	37.2	100
Metallurgiche	40.7	59.3	100	47.2	52.8	100	66.5	33.5	100	55.6	44.4	100
Meccaniche	49.5	50.5	100	56.9	43.1	100	65.5	34.5	100	58.7	41.3	100
Elettriche	59.8	40.2	100	64.3	35.7	100	77.4	22.6	100	74.0	26.0	100
Tessili	24.6	75.4	100	28.9	71.1	100	38.5	61.5	100	31.6	68.4	100
Abbigl. e Cuoio	27.5	72.5	100	19.5	80.5	100	46.2	53.5	100	29.1	70.9	100
Legname e affini	26.6	73.4	100	44.9	55.1	100	57.1	42.9	100	42.0	58.0	100
Agglom. Laterizi	37.7	62.3	100	47.0	53.0	100	62.1	37.9	100	48.8	51.2	100
Ceramica e vetro } Alimentari e Pesca }	32.9	67.1	100	36.7	63.3	100	48.3	51.7	100	40.4	59.6	100
Carta e Poligrafici	50.5	49.5	100	57.6	42.4	100	63.5	36.5	100	57.3	42.7	100
Commercio	30.3	69.7	100	33.6	66.4	100	44.5	55.5	100	34.6	65.4	100
In Complesso	37.1	62.9	100	42.4	57.6	100	58.5	41.5	100	47.67	52.33	100

pendenza per le quali la schiera degli uomini nuovi irrompe nel ceto superiore degli imprenditori, per poi dilagare nelle altre generazioni per tutto il complesso.

La tabella 2 riporta, in cifre assolute e proporzionali, questa distribuzione combinata. Si rileverà da questa tabella che la percentuale dei "non laureati" è massima per quei gruppi industriali per i quali susiste un numeroso artigianato (abbigliamento, tessili, alimentari) e che assume i valori più piccoli per quei gruppi di attività in cui questo artigianato ha proporzioni minime (industrie elettriche, estrattive).

Se si raffronta poi la frequena dei "non laureati" dei vari gruppi, con quella degli esercizi artigianali rispetto al totale degli esercizi, o degli addetti in esercizi artigianali rispetto al totale degli addetti, secondo le risultanze del censimento industriale del 1937-40[12] si notera una forte concordanza positiva tra la prima e le altre due serie:

TABELLA 3

Frequenza dei Consiglieri non Laureati nelle maggiori società industriali, e frequenza degli esercizi artigianali e degli artigiani nelle rispettive categorie industriali

Categorie industriali	consiglieri non laureati	artigiani e dipendenti in percento del totale	Esercizi artigianali
Elettriche	26.0	0.0	0.0
Estrattive	37.2	0.0	3.3
Meccaniche	41.3	20.3	95.0
Carta e Poligrafiche	42.7	9.2	50.3
Metallurgiche	44.4	0.0	0.0
Agglomeranti, Laterizi Ceramica Vetro	56.1	14.2	68.3
Legno e affini	58.0	63.5	93.8
Alimentari e Pesca	59.6	32.1	58.8
Tessili	68.4	5.8	74.2
Abbigliamento, Cuoio, pelli	70.9	74.9	98.2

Tutto questo ci fa intendere che le linee di maggior pendenza per le quali scorre il flusso vivificatore di rinnovazione sociale passano per quelle attività in cui, anche storicamente, l'azienda familiare si svilupa nella piccola e poi nella media e grande impresa.

La scuola superiore in tutto questo movimento non ha dunque che una parte secondaria. La maggioranza dei giovani che ne superano il vaglio non raggiunge poi una posizione sociale parificabile per prestigio e per reddito a quella che raggiungono, forse più facilmente, coloro che superando il vaglio della pratica professionale salgono alla direzione di impresse industriali e commerciali, di media o anche di piccola importanza.

---

[12] I dati sono ripresi dal Vol. I, Industria, Parte I, Esercizi, Addetti, forza motrice.

L'esperienza fa anche temere che la proporzione dei laureati che sostano e languono in basso sia maggiore tra coloro che provengono dal ceto dei lavoratori manuali, che non quelli che provengono dal ceto medio.

Poichè è credibile che molti, per funesto equivoco, tendano all'università o vi siano avviati non per le loro attitudini o tendenze spirituali, ma solo perchè si vede nell'università un mezzo di elevazione economica e sociale.

Per il presente almeno, se si vuol veramente favorire il processo di circolazione dal basso all'alto, in modo che all'alto salgano i più meritevoli conviene sopratutto favorire quelle forze che piu occultamente, ma pur tanto efficacemente, sono già in azione.

Perciò, nel campo dell'istruzione, le maggiori cure per raggiungere questo fine non vanno indirizzate verso la scuola superiore, ma bensi verso la scuola media professionale e specialmente verso quella più bassa, affinchè questa istruzione venga a favorire sempre di più coloro che per il filone della pratica, e dopo il vaglio della selezione professionale, debbono vincere la concorrenza degli altri più favoriti dall'ambiente.

Concludiamo con l'augurio che queste ricerche sul volume della circolazione sociale, e sulle vie principali che essa segue, siano ripresi ed approfondite.

Certo come dicevamo, la Statistica ufficiale non sembra molto adatta ad effettuare rilievi de carattere così delicato. Al contrario questi rilievi possono essere compiuti, ci sembra, da quelle organizzazioni che hanno per scopo l'effettuazione di inchieste col metodo dei campioni.

Il nostro Istituto potrebbe prendere l'iniziativa per favorire queste richerche e convogliarne l'effettuazione in un quadro uniforme,

Résumé

It is desirable that representative inquiries should be made, so as to obtain as complete information as possible with regard to:

a) the quantitative importance of social mobility;
b) the lines it follows;
c) its degree of selection.

The International Statistical Institute might develop a very useful action with the object of promoting these inquiries in an organic setting and of reducing the incomparability caused in such studies by the complexity of the phenomenon and the ambiguous character of various nations.

The author exposes certain results of a statistical study which he made in this field. These results are of particular interest in regard to the item *b* mentioned above.

In Italy the number of graduates whose fathers belong to the labouring class amounts to about 12%; the same proportion has been found in other countries of Central and Western Europe.

It is probable that almost all sons of workmen who get higher education (different from the others) will occupy posts for which a university degree is required. So it may be presumed that in these functions the influx of new men is rather strong. But the more important question is to know the proportion of the new men occupying important functions for which no university degree is required.

The author has endeavoured to get an idea of the situation in this respect for a group of industrial employers more precisely the members of the board of directors of limited liability companies, the social capital of which was at least a million lire in 1939.

The source of data used for this purpose permitted to distinguish those who had obtained a university degree from those who had not and those who had a nobiliary title from the others. By conjecture, this information has permitted to conclude that among the people under consideration, the proportion of new men would be considerably higher than 27%.

The rate of persons without university degree decreases with increasing social capital. For each size of firm (according to the social capital), the proportion of persons without university degree is highest in the clothing and leather industries; this proportion is likewise rather high in the textile industry; it falls in metallurgy and the mechanical industry, and reaches a minimum in mining.

Thus it may be concluded that in the class of leading industrialists the rate of new men is higher for those industries having their basis in the small scale home-industry.

It is unfortunate that only such limited sources of data are available for research work in this field. Official statistical agencies do not seem to be the most suitable for making the necessary enquiries, as the questions to be asked are very delicate. This might be more satisfactorily undertaken by organisations engaged in public opinion research, and sample surveys in general.

The International Statistical Institute might take the initiative of promoting such inquiries on a uniform basis.

Résumé

Il semble désirable que des enquêtes représentatives soient faites de façon à obtenir des renseignements aussi complets que possible sur:

a) l'importance quantitative du mouvement de capillarité sociale,

b) Les voies qu'il suit,

c) son degré de sélection.

L'Institut International de Statistique pourrait développer une action très utile, ayant pour but de favoriser ces enquêtes dans un cadre organique et de réduire l'incomparabilité que la complexité du phénomène et le caractère équivoque des définitions causent dans de telles études.

L'auteur expose certains résultats tirés d'une enquête qu'il a faite dans ce domaine. Ces résultats offrent de l'intérêt surtout pour ce qui se rapporte à l'importance des diverses voies suivies par les mouvements de capillarité sociale.

En Italie, les lauréats ayant un père ouvrier (ou de situation similaire) ont à peu près, parmi tous les lauréats, la proportion de 12%: la même proportion qu'on a relevée en d'autres pays de l'Europe Centrale ou Occidentale.

Il est vraisemblable que les fils d'ouvriers qui fréquentent les écoles supérieures sont, presque tous (à la différence des autres) acheminés vers les professions ou les emplois pour lesquels un diplôme universitaire est nécessaire. Pour cela on doit supposer que dans ces professions l'affluence des "hommes nouveaux" soit assez forte. Mais la question la plus importante est de savoir quelle est la proportion des "hommes nouveaux" exerçant des fonctions directives dans les activités pour lesquelles un diplôme universitaire n'est pas nécessaire.

L'auteur a tâché d'approcher la réalité pour un groupe d'entrepreneurs de l'industrie, plus précisément les membres des conseils d'administration des sociétés par actions dont le capital social, en 1939, n'était pas inférieur à un million de lires.

La source utilisée à cet effet permettait de distinguer parmi eux, ceux qui avaient obtenu un diplôme universitaire, de ceux qui ne l'avaient pas, et ceux qui possédaient un titre nobiliaire, des autres. Par conjectures, ces données ont permis d'affirmer que dans le total des unités considérées, la proportion des "hommes nouveaux" serait sensiblement supérieure à 27%.

On constate que la proportion des personnes sans diplôme universitaire diminue d'autant plus que le capital social est élevé.. Dans toutes classes de capital social, la proportion des personnes qui n'ont pas de diplôme universitaire est la plus forte dans l'industrie du vêtement et du cuir; cette proportion est de même assez forte dans l'industrie textile; elle se réduit dans la métallurgie et la mécanique, et elle atteint le minimum dans l'industrie minière et électrique.

On peut conclure de cela que, dans la classe de grands entrepreneurs industriels, la proportion des "hommes nouveaux" s'élève d'autant plus que la grande industrie a, à sa base, la petite industrie à caractère familier.

Il est pénible que les savants qui se consacrent à ces recherches soient obligés d'utiliser des sources si limitées.

D'ailleurs, la statistique officielle ne semble pas la plus adaptée à effectuer les relevés nécessaires, dont les questions à poser sont très délicates. Ces relevés peuvent mieux être faits, nous semble-t-il, par les organisations de recherches de l'opinion publique, et, en général, d'enquêtes par la méthode des sondages.

L'Institut International de Statistique pourrait prendre l'initiative de favoriser ces recherches et d'en diriger l'accomplissement suivant un cadre uniforme.

COST OF LIVING, SALARIES, CONSUMPTION AND THE SITUATION OF THE ITALIAN MIDDLE CLASSES

by Guglielmo Tagliacarne
Italian General Federation of Commerce

I

1. The Increase in Living Costs

During the first years of the war it was possible to keep the cost of living within reasonable limits; but towards the end of the conflict prices began to soar rapidly and after the armistice they increased even more rapidly. In fact, by the end of 1943 the index of the cost of living was 4½ times that of before the war; at the end of the following year it had risen to 12 times the prewar; by the end of 1945 it was up to 25 times; and at the end of 1946 the index was 39 times that of the period preceding the conflict. Now (May 1947), prices are 50 times the 1938 level.

Not all the prices of the various items that go to make up a family budget have increased to the same extent. The cost of food has increased 67 times, of heat and light 22 times, while sundry expenses have increased 24 times. Rents, however, were as of May 1947 only 2½ times those of 1938.[1]

If total consumption had remained unchanged in both quantity and quality, compared to 1938, the different increases in the prices of the various items of the family budget would have increased the percentage of the budget going for food, raising it from 51% (average for 1938) to 67% (May 1947) of the total budget. This means that, while in 1938 expenses for food absorbed little more than half of the total amount spent by a working family, today two-thirds of all family expenditure would be required to make the same purchases. Clothing expenses, from requiring 17% (1938) of the total expenditures, would now (May 1947) require 26%. The part going for rent, which had been approximately equal to that fixed for clothing, *i.e.*, 16% of the total, appears to have fallen to a negligible consideration, to less than 1% of the total.[2]

---

[1] In Italy various series of index numbers are being calculated to express the cost of living, with results that differ notably one from another. This is due not only to the different methods employed in the calculations but to the irregularity of business transactions and the difficulty in obtaining reliable statistical reports especially because of the many products and articles bought on the black market. The data referred to in this text are based on an average of three indices obtained from the followng three sources: (1) Economic Service of the Montecatini Co., (2) Technical Committee of the Edison Group, (3) National Index of the Central Institute of Statistics.

[2] These figures are based on the cost of living indices drawn up by the Edison Group.

Expenses for food, however, have absorbed much more than the above-mentioned quota of 67%. As was pointed out, this percentage is based on the hypothesis that total consumption in both quantity and quality has remained the same as before the war. But, actually, the sharp drop in incomes did not allow the population to maintain the prewar standard of living; it was forced to make sacrifices with consequent restrictions in its habitual consumption.

Obviously, reductions can be made more easily in the sectors least indispensable, while it is especially impossible to reduce in any great measure the consumption of foodstuffs. This was the case in Italy where food consumption was already quite low even before the war. Expenditures for food today, therefore, demand not 67% but at least 75% of the family budget. Clothing expenses have decreased to the point where they now probably absorb not more than 15%, far less than the above-mentioned 26% required had 1938 consumption been maintained.

2. THE INCREASE IN PUBLIC SERVICE COSTS AND TAXES

It is interesting to examine the changes in cost of the most important public services and certain articles and products as compared to the price changes of food and clothing. The indices of some of these items are as follows:

TABLE 1

Representative Costs, May 1947—Milan or Rome
(1938 cost = 1)

Gas for cooking—Milan		15
Electricity for lighting (with meter)—Milan		7
Water—Rome		4
Daily newspaper		33
Subscription for radio broadcasts		12
Doctors' fees—Milan		20
Postage stamps		12
Railroad fare (3rd class, 100 kilometers)		8
Trolley fare (non-holiday)—Milan		14
Taxi fare—Rome		35
Haircut—Rome		15
Public telephone		10
Movies—Rome		20
Theatre—Rome		10
Tobacco (controlled price)		14
Direct taxes		8
Average expenses for food	67	
Average expenses for clothing	69	
Over-all cost of living	50	

Sources: Technical Committee of the Edison Group; UNRRA; Central Institute of Statistics.

It is readily seen that the increases in service prices are much smaller than those in prices of food, clothing, and the over-all cost of living.

3. Official Prices and Black Market Prices

Though the Government fixes official prices for a large number of rationed items, in practice large additional quantities of these items are sold at much higher prices on the black market.

The following examples for a group of food products which are consumed on a large scale in Italy give an idea of the difference existing between the two prices:

Table 2

Comparison of Official and "Black Market" Retail Prices
Rome, May-June 1947

	Lire per kilogram	
	Official	"Black Market"
Bread	34	195
Pasta (spaghetti, etc.)	53	278
Rice	65	242
Bacon	510	950
Olive oil	355	838
Sugar	165	813
Maize flour	27	70

Source: *Rivista del Commercio*, 1947, No. 3.

4. Prices Based on Caloric Values

It is also interesting to examine the prices of food products in terms of their caloric values. From this point of view one can divide food products into low-priced and luxury foods, the first category including cereals, the second including especially meat. Butter also is relatively expensive as compared to olive oil, which can be used in its place.

Table 3

Cost of 1,000 Calories of Various Foodstuffs
Free market, Rome—June 1947

	Lire
Wheat flour	52.24
Maize flour	19.14
Home-made bread	71.32
Pasta	73.57
Rice	68.70
Beans	60.00
Lentils	56.46
Bacon	107.32
Butter	130.37
Olive oil	92.96
Sugar	108.17
Meat	1,000.00

Source: Calculated by the author on the basis of the caloric values of these products as estimated by the Institute of Nutrition of the National Research Council (Rome).

A diet consisting only of *polenta* (a pudding made with maize flour) would be anything but health since *polenta* does not contain all the nutrition elements necessary to the human body; but apart from this one could say that 57.42 lire a day (800 grams of maize flour) would suffice to give an active man the 3,000 calories indispensable to him (expenses for cooking and salt should be added).

For similar reasons a diet of meat alone is no more advisable than one of only *polenta.* Furthermore, it is expensive indeed. A diet consisting exclusively of meat would mean a daily expense of 3,000 lire to provide the 3,000 calories required by an active man. In other words, a meat diet is 52 times as expensive a diet of *polenta.*

5. COMPARISON BETWEEN THE FIRST AND SECOND WORLD WARS

A comparison between prices after this war and those after World War I can be drawn by examining the indices of the cost of living in Milan; let us begin for World War I with January 1919 and for the second with July 1945.

Immediately after the end of the first conflict the indices of the cost of living decreased and by July 1919 they were down 23%. Later, prices began to rise so that by December 1919, one year after the end of the war, the cost of living was back to its January 1919 level.

In this second postwar period prices have followed a somewhat different pattern. Principally, there was no time following the conflict during which prices actually dropped; instead, there has been a continuous increase, an increase which was moderate during the first postwar year but which subsequently turned into a situation in which prices soared rapidly.

In both the first and second postwar periods, the most serious increases in the cost of living started in the second year after the end of hostilities. Obviously, this comparison between prices then and now can be presented only as indicative since there appears to be no valid reason for believing that prices during the two postwar periods should necessarily follow the same pattern.

The greatest difference in the trend of prices during the two periods is to be found in the extent of the increases.

TABLE 4

Indices of the Cost of Living in Milan

World War I (1914 = 100)*			World War II (1938 = 100)**		
1919	January	353	1945	July	1,802
	April	331		October	2,415
	July	271	1946	January	2,433
	August	288		February	2,429
	September	302		March	2,586
	October	323		April	2,536
	November	348		May	2,529
	December	352		June	2,609
1920	January	378		July	2,652
	February	382		August	2,817
	March	374		September	2,804
	April	392		October	3,123
	May	438		November	3,430
	June	448		December	3,868
	July	441	1947	January	3,919
	August	449		February	3,842
	September	470		March	4,245
	October	489		April	4,513
	November	511		May	4,930
1921	January	571			
	February	566			
	March	568			
	April	578			
	May	578			
	June	506			
	July	494			

* Source: *Annuario storico-statistico* of the Municipality of Milan.

** Source: Economic Service of the Montecatini Co.

Let us extend our comparison to the various items included in a family budget. Two years after the end of each world war the various items of a family budget were represented by the following numerical indices:

	First World War December 1920 (1914 = 100)[3]	Second World War May 1947 (1938 = 100)[4]
Food	535	6,682
Clothing	782	6,932
Rent	139	251
Heat and light	886	2,228
Sundry expenses	604	2,438

[3] Source: As in Table 4.

[4] Source: Average of three indices, Montecatini Co., Edison Group, and Central Institute of Statistics. In Table 4 only the indices drawn up by the Montecatini Co. were used because those of the Edison Group and the Central Institute of Statistics begin at a more recent date.

II

6. THE REDUCTION IN THE PURCHASING POWER OF WAGES

As was to be expected, wages also tended to increase in an effort to keep pace with the rise in prices and cost of living.

The first step in this study is to compare the rise in wages with that in the cost of living. In the course of this survey we shall find that the situation differs for the various categories of wage-earners and we shall have to distinguish between the various cases, limiting ourselves, however, to typical examples so that we shall not lose ourselves in a long account examining all occupations and the hierarchy of categories.

For the comparison which we are about to make we shall adopt as our base the average of the three indices of the cost of living which we have used before; according to this base the cost of living in May 1947 was about 50 times that of 1938.

As of this month the wages of an unskilled laborer, instead of the 50-fold increase required to keep pace with the increase in the cost of living, had increased only 36 times; the wages of a skilled workman increased 26 times, the salary of a second class employee 19 times, of a first class technician 13 times, and of an executive 12 times (Source: Economic Service of the Montecatini Co.).

Some important observations are immediately obvious:

a) the increase in the cost of living has been greater than the increase in wages;

b) this divergency becomes greater as one moves up the scale from the grades of unskilled to those of skilled labor, salaried personnel, and finally executive class;

c) the percentage increase in the wages of unskilled laborers has been three times that of executives;

d) real wages, *i.e.*, wages in terms of their purchasing power, have decreased, compared with 1938, 24% for unskilled labor, 46% for skilled, 60% for second class employees, 70% for technical first class employees, and 72% for executives.

While, therefore, the unskilled laborer has succeeded in keeping his standard of living from being very greatly reduced the same cannot be said of workmen and office workers in the higher categories. The phenomena resulting from this notable shrinking of economic distances continue to be far-reaching and resultant of consequences of the greatest importance.

The above-mentioned trend in wages paid by private firms is even more marked for state and public service employees (see figure 1).

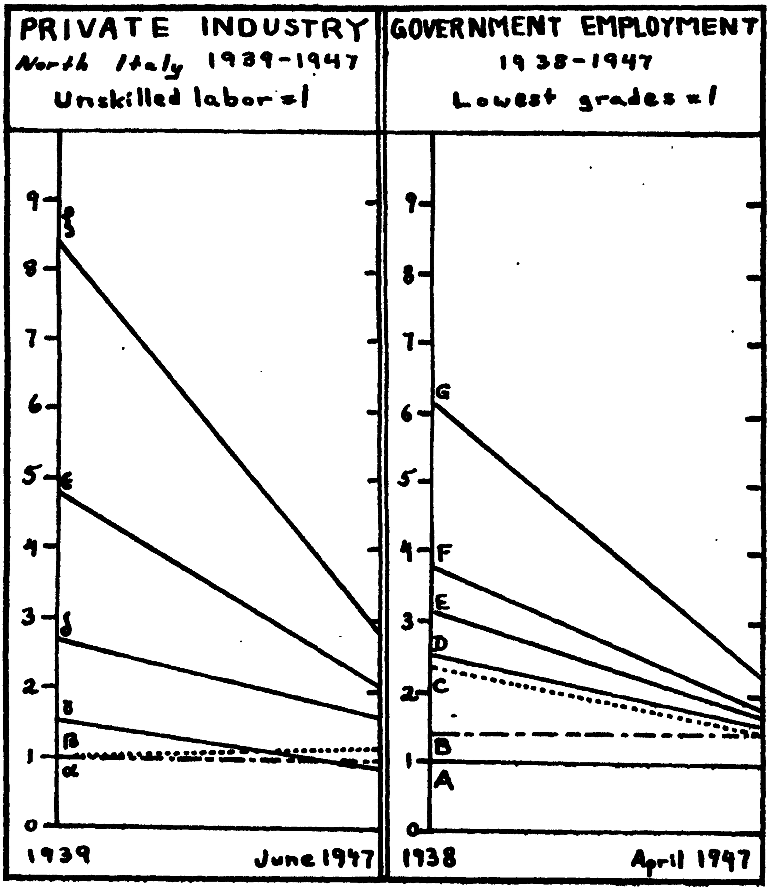


Figure 1

This "fan" of wages, moderatey spread out before the war, has been contracting more and more with the worsening inflationary situation and the rise in the cost of living, until it threatens to close completely.

7. Effects of the Equalizing of Wages

The trend toward equalizing of wages among the various grades of labor has resulted in a real displacement of the middle and intellectual classes. Already the effects of this phenomenon can be seen, effects partially favorable and partially unfavorable.

National income, it will be remembered, has decreased 35 to 40% as compared to 1938. Even then it was certainly not high. If this decrease had been uniformly distributed among all classes, the conditions of the lowest categories of workers would quickly have fallen below the starvation level, and their health, resistance, and capacity for work would have suffered greatly with serious consequences for the whole of society. Instead, the leveling of salaries suffered by the upper incomes was not, in general, drastic enough to undermine their health or their working powers. From this point of view no noticeable damage is felt by society, while under the first case the consequences would have been far-reaching and serious.

As a result of the equalizing of wages, consumption becomes modified at the expense of luxury goods or at least of those goods which are of less universal demand, being directed instead to those more common and standardized goods that are manufactured at a lower cost.

Another effect of the equalizing trend will be that the numbers in the intellectual classes will tend to decrease, partly because their years of preparation are not being recompensed and partly because they lack the means for continuing their studies and for dedicating the years necessary to the training for a profession.

The worsening of the living conditions of the middle classes (which include white collar workers, the large majority of the independent professionals, artisans, merchants, small and medium land-owners) as a result of the drastic reduction of real incomes has forced the heads of families to resort to the supplementary work of members of their families.

The situation also has its repercussions on the relations between state employees and the State, of the quantity and quality of the work of the employees and consequently on the complex problem of bureaucracy. State and public service employees receive salaries which are inadequate in terms of either the cost of living or their social standards. As a result they take on positions hostile to or in contradiction with the authority of the State; they are often forced to accept supplementary work in their spare time.

The lowering of real wages to the level of that of the most modest and most unskilled worker cannot but have a harmful effect on the sense of hierarchical discipline between the various grades of workers.

Finally, one must consider the effect of this wage equalization trend on the accumulation of savings. It is well known that in normal times the middle classes are the largest contributors to the various forms of savings and therefore supply a large part of the funds which eventually, through banks, etc., find their way into productive activity in industry and commerce. The grave economic situation in which these classes now find themselves certainly does not allow them to continue these habits. On the other hand the more favored categories of workers have not yet acquired the savings habits. It should also be pointed out that in times such as those we are examining, characterized by rapid inflation and an even more rapid increase in prices, saving habits are certainly not encouraged.

III

8. AVAILABILITY OF FOOD

Following the exceptionally poor 1945 harvest in cereals and other agricultural products, the situation in 1946 improved notably allowing, with the help of a substantial UNRRA contribution, a moderate increase in food consumption. Italy's dependence on foreign countries for food had been gradually decreasing before the war; from 16% in 1922 (79 kilograms per peon) rsit had been reduced to 3.3% (19 kilos per person) in 1938 without any reduction in the total amount of food available for consumption, which in those years amounted to 576 kilos of food products per person.[5]

In 1946 the amount of imported food products totaled 1,793,073 quintals, equal to 39 kilos per inhabitant (exactly double the amount imported in 1938) which means that in 1946 Italy brought in 9% of her food from abroad.

Despite the fact that the percent of foodstuffs imported had increased over the prewar period, the quantity of food available to each inhabitant diminished. In the 1933-38 period their diet contained nutritive elements yielding 2,500 calories daily while in 1945-46, even with increased importation they received only 1,520 calories. In 1946-47 the availability of calories received a substantial boost; and, although definite figures for that year are not yet available, it can be estimated that the average per capita intake of calories increased to more than 2,000.

But although it has increased in the last year, the average diet in Italy is still 20% below the prewar level. What makes matters worse is the fact that its nutrition value has become poorer since it is largely composed of foodstuffs of vegetable origin. In fact, in the years 1933-38, 78% of the Italian diet was of vegetable origin and 22% of animal; in 1945-46 animal products had fallen to 17%;[6] in 1946-47 it was probably up to 20%.

9. RATIONING

Confirmation of the improved availability of foodstuffs in 1946, especially after the good harvest of that year, came with the increase in rations obtainable by ration card. From a total of 747 calories per day per normal consumer in the first quarter of 1946, ration availabilities climbed to 958 in the last quarter of that year. That this increase was not of an entirely seasonal nature is shown by the fact that in the first months of 1947 there was a substantial increase over the preceding year. The rations distributed by ration card during the first five months of the current year had to be reduced because of the shortage of wheat supplied, but despite this they furnished 851 calories per day as against 729 for the corresponding period of 1946.[7]

---

[5] G. Tagliacarne, *Le vie di rifornimento dell'Italia* (The Supply Roads of Italy), Confederazione dei Commercianti, Roma, 1940.

[6] UNRRA, Italian Mission, *Survey of Italy's Economy*, 1946.

The average daily ration distributed by ration card in 1946 contained a total of 839 calories, about one-third of the normal consumption on the basis of prewar averages. The ration for the first five months of 1947, consisting of 851 calories, satisfied no more than 40% of the total reduced food consumption. The other 60% of food needs were being met by products bought on the free market (nonrationed products) or on the black market (rationed products).

Bread, a fundamental element in the Italian diet, supplies 73% of the total caloric value assured through the ration card. And yet the bread ration is only 235 grams daily per capita, which is the lowest of any country in Europe. Several months ago this ration was further reduced to 200 grams, the other 35 grams being replaced with distributions of maize flour.

Products used for making soup (pasta, rice, legumes, flour) contribute 17% of the caloric value distributed by ration card; fats represent only 5%, sugar 4%, and meat is insignificant (average 1946).

TABLE 5

Average Daily Caloric Value of Rations Distributed in 8 Cities, For Normal Consumer

	Bread	Soup products, legumes	Fats, meat, milk products	Sugar	Total
Average of first 5 months, 1946	563	98	47	21	729
Average of first 5 months, 1947	571	186	54	40	851

Source: Data from the Central Institute of Statistics. Our calculation for the average of eight cities.

10. MEAT CONSUMPTION

Statistics of taxes on meat sales and on animal slaughter give us an idea of the changes that took place in the consumption of meat in 1946 as compared to the prewar period. According to our calculations using the data of the above-mentioned sources, meat consumption decreased from 31.10 kilos per capita in 1938 to 22.78 in 1946, a drop of 27%. The most important decreases were in the consumption of veal and pork, the most expensive types of meats, while there has instead been a strong increase in the consumption of horse meat. The lowering of quality level shows up clearly in the following table in the comparison of the percentages:

---

[7] These figures are an average for 8 principal cities. They include only rations distributed to "normal" adults and exclude the supplementary rations given to children, old people, and laborers doing heavy work.

TABLE 6

Composition of Meat Consumption

	1938		1946	
	Kilos per capita	% of total	Kilos per capita	% of total
Veal	10.42	33.51	6.48	28.45
Other beef	12.56	40.38	9.36	41.09
Pork	5.56	17.88	3.30	14.49
Poultry	2.18	7.01	2.04	8.95
Horse	0.38	1.22	1.60	7.02
	31.10	100.00	22.78	100.00

Note: These figures are based on sales taxes on meat; figures for 1946 are for the first half year multiplied by two.

11. TOBACCO CONSUMPTION

The reduction in the consumption of cereals, meats, and food in general has been imposed by the shortage of these products; it will be interesting to examine now the trend in the consumption of other products and luxury expenditures.

An important index is tobacco. Per capita consumption decreased from 653 grams in 1938 to 547 grams in 1946, but it is now increasing appreciably month by month, and during the first three months of 1947 it had surpassed the prewar level. In fact during these first three months it was 673 grams per capita as compared to 494 grams during the corresponding period of 1938.

These statistics are the official ones. It must be pointed out that, as everyone knows, besides the national tobacco, which is controlled by a State monopoly and rationed, there is a flourishing black market in foreign tabacco. Although there is no way of calculating the amount thus sold, it is certain that it is sufficient to bring today's per capita tobacco consumption to a level at least 50% higher than before the war.

12. EXPENSES FOR AMUSEMENT

An analogous trend has been taking place in expenditures for theaters, movies, and gambling. The latter has reached such large proportions in the postwar period as to surpass the combined total spent on the theaters and sport events. But the largest share of the expenses for amusement still goes to the movies, which in 1946 absorbed three-fourths of the total spent by the Italian population on entertainment as compared to 71% in 1938. For this type of entertainment the population of Rome in 1946 spent an average of 1,150 lire per capita to see 21 pictures. For all of Italy it has been calculated that each Italian in 1946 went to the movies or the theater about 12 times as compared to 10 times in 1938.

The figures for the first months of 1947, though not yet definitely known, indicate a continuous increase not only in expenditures, which is largely due to the increase in the prices of tickets, but also in the number of performances attended.

13. THE DISPLACEMENT OF THE MIDDLE CLASSES

Although an increase in certain luxury expenditures is shown in the two preceding chapters, one must not get the impression that this demonstrates an improvement in the general level of living. This type of increase in expenses is explained rather by changes in the relative positions of the various social classes of the population and by changes in customs and habits.

Every war, with the inflation and social upheavals that follow in its wake, creates classes which are particularly favored by the events. In general these classes are composed of enterprisers, and in particular of persons who deal in military supplies and in items which are especially in demand, such as food, durable goods, etc. At the top of this large group of persons favored by the events of the war and the postwar inflation, one finds numerous profiteers and speculators. These are the groups that help increase the consumption of luxuries. Other social groups which have been unaffected or affected only slightly by the drop in real incomes (farmers and certain classes of workmen) also contribute to this increase.

In contrast to this, the vast middle classes, which in Italy are estimated at 10 million persons (20% of the population) suffered much from the drop in their incomes, which remained far below the increase in prices and cost of living.

Special surveys[8] on the condition of these classes in the postwar period indicated the great sacrifices undergone to hold on the way of life which even under adverse conditions distinguishes them from the other strata of the population. They fell back on accumulated savings and sold objects of art, jewels, furniture; they pared all superfluous expenditures from their budget and lowered ther standards of consumption; finally they were forced to send the "idle" members of the family out to look for gainful occupations.

---

[8] See, for example, the survey presented by Prof. Ugo Giusti at the meeting held in April 1947 by the Italian Society for Demography and Statistics.

It is our conviction, as we mentioned before, that this phenomenon we have surveyed is a passing one, the result of a rapid inflation and of the sharp cut in the national income, a phenomenon therefore necessary under the present exigencies but destined to return to the normal differentiations of the prewar period as soon as the lira is stabilized, industrial development resumes its course, and national wealth increases.

Résumé

Le coût de la vie en Italie est considérablement accru après la guerre, notamment depuis la fin de 1945. En comparaison avec l'augmentation des prix après la première guerre mondiale, nous pouvons affirmer que dans ce deuxième après-guerre le coût de la vie a augmenté à peu près dix fois plus (en juillet 1947, 50 fois les prix de 1938) qu'il avait augmenté après la première guerre (à la fin de 1920, 5 fois les prix de 1914). Les salaires, et notamment les appointements des employés, n'ont pas été augmentés dans la même proportion. Les ouvriers non qualifiés ont vu leur salaire s'accroître au moins 40 fois, c'est-à-dire le salaire réel s'est réduit de 20 pour cent environ. Mais les appointements des fonctionnaires plus élévés ont vu leur salaire réel se restreindre de 60 pour cent (et quelquefois plus) de ce qu'il était avant la guerre.

Cette tendance à l'égalisation des rétributions acquises aux dépenses des employés a produit un véritable déclassement des classes moyennes. Les individus de ces classes de la société ont dû se sacrifier, c'est-à-dire réduire les dépenses de l'alimentation aussi bien que celles des vêtements et de l'instruction, et enfin ils ont dû avoir recours dans une large mesure aux épargnes accumulées dans le passé. Les femmes et les enfants ont dû chercher un emploi pour compenser l'insuffisance du salaire du chef de famille. Les conséquences ne sont pas négligeables. Il y a une réduction du montant de nouvelles épargnes parce que les nouveaux riches n'ont pas acquis l'habitude et l'esprit pour s'imposer de tels sacrifices; les employés ne sont pas encouragés à étudier, à perfectionner leur préparation, puisque leurs fatigues et leurs sacrifices ne sont pas suffisamment récompensés. D'autre part, le sentiment de la hiérarchie est sérieusement atteint.

Il se peut que cette situation du nivellement des salaires soit une nécessité du temps présent, très pénible pour tous; mais s'il n'en était

pas ainsi, nous devrions nous attendre à un abaissement dangereux du niveau intellectuel général, de l'esprit d'épargneet enfin une réduction relative des classes moyennes, très redoutable, car elles sont des éléments d'équilibre, de stabilité et de paix dans la société nationale.

Discussion

Mr. Agapitides added a few words concerning the situation of salaried people in Greece. Indices which he has prepared show that during the war, *e.g.*, real salaries of public officials decreased by April 1944 to 7% of their prewar level.

Mr. Agapitides described the effect of stabilization of the national currency after liberation, when real salaries first increased to 70% of prewar and then dropped again with a new inflation of the national currency. Repeated attempts at stabilization have only been followed by inflation.

SOCIAL STATISTICS SECTION

Monday, September 15, at 9:30 a.m.

Public Opinion Research

Chairman :

P. J. Idenburg

LES PROGRES RECENTS DE L'ETUDE DES OPINIONS EN EUROPE

par Jean Stoetzel

Le beau travail qui vous a été présenté samedi dernier par le professeur Frederick F. Stephan sur "l'histoire de l'utilisation des procédés modernes d'échantillonage", pourrait servir de modèle à une histoire des enquêtes d'opinion. Les curieux n'ont certainement pas fini de trouver des ancêtres de plus en plus reculés aux enquêtes de *Gallup* et de *Fortune*. *Emil Hurja* a cité deux enquêtes américaines précoces, réalisées en 1824. Mais le comte de *Lavalette* en France, quelque vingt cinq ans plus tôt, et *Daniel Defoe*, en Angleterre, plus d'un siècle auparavant, avaient déjà mis sur pied des organisations pour le sondage des opinions.

Cependant, en dépit de l'attrait de l'anecdote et des satisfactions que peut procurer la préhistoire, il reste légitime d'affirmer que l'étude proprement scientifique des opinions dans le monde n'a pas une histoire de plus de vingt cinq ou trente ans. Si l'on se borne à ce qui est vraiment l'essentiel, il faut faire remonter à la décadé 1920-1930 le premier ensemble convergent de recherches sur les opinions. L'apparition des premiers sondages de l'opinion publique ne dépasse guère 1935. Encore faut-il ajouter que les premières réalisations sont toutes américaines. J'aurai donc à vous parler ici du plus récent parmi le récent.

En 1939, à la veille de la guerre, l'Europe n'avait encore guère manifesté d'intérêt pratique pour l'étude des opinions. Si l'on met à part les recherches du professeur *André Siegfried*, basées sur l'analyse des résultats des élections françaises dans différentes provinces, recherches remarquable par la précision de la méthode et par la finesse des hypothèses qui les sous-tendaient, mais cependant recherches d'archives, recherches passives par conséquent, et de plus extrêmement limitées dans le champ des opinions, on ne trouve guère de preoccupation de ce genre dans les milieux universitaires. Deux organisations de sondages seulement existent, et de fraîche date, une en Grande-Bretagne, et une en France.

En 1947, nous rencontrons au moins quinze organisations de sondages en Europe, distribuées dans douze pays, à savoir: Allemagne (deux organisations, au moins); Belgique, Danemark, Finlande, France (deux organisations), Grande-Bretagne, Hongrie, Italie, Norvège, Pays-Bas (deux organisations), Suède, Tchécoslovaquie. Encore faut-il ajouter qu'il existe une deuxième organisation en Norvège, et une troisième en Allemagne, sur lesquelles je n'ai pu recueillir de documents precis; et qu'il y a aussi un second institut en Belgique, et un en Suisse, ces deux derniers cependant ne paraissent pas avoir dépassé la période expérimentale. Je devrai m'abstenir de parler de ces quatre organisations, qui porteraient à dix-neuf le nombre pour l'Europe.

Ainsi, l'étude européenne des opinions est toute récente: ceux instituts fonctionnaient avant 1940; deux apparaissent entre 1940 et 1944; et onze se développent depuis 1944. Il est juste d'ajouter que parmi ces onze instituts, deux au moins avaient déjà jeté leurs premières bases avant 1944. La création et le fonctionnement des instituts d'opinion en Europe n'ont pas manque, au moins au début, d'être accueillis avec indifférence, scepticisme, voire même d'être l'objet de vives critiques. Il est assez remarquable que ce sont surtout des économistes et des personnes appartenant au milieu des affaires, qui ont pris surtout l'initiative de créer de telles organisations. C'est ce qui s'est passé dans huit cas sur quinze. Les sept autres instituts ont été fondés respectivement par: deux statisticiens, deux psychologues, un sociologue, un militaire, et un ancien consul.

Il semble que l'on assiste actuellement en Europe à une acceptation de plus facile de la nouvelle branche de recherche. L'âge du directeur à la date de la création de chaque nouvelle organisation a tendance à s'élever; ses autres fonctions, antérieures ou simultanées, évoluent dans un sens bien défini; la forme même des organisations change. Pour la période antérieure à 1940, les hommes qui ont créé les deux premières organisations de sondages sont âgés en moyenne de 30 ans à la date de la fondation; entre 1940 et 1944, les nouveaux dirigeants ont 35 ans en moyenne; depuis 1944, l'âge moyen des directeurs des onze plus récents instituts sera, toujours aux dates respectives de leur fondation, de 40 ans environ. De même, tandis qu'avant 1944, on trouve seulement un universitaire parmi les quatre directeurs, depuis 1944, les onze nouveaux instituts ont à leurs tête cinq universitaires et trois autres fonctionnaires, plus de la moitié par conséquent. Enfin, toutes les organisations antérieures à 1944 sont commerciales, alors que, parmi les onze nouvelles, cinq seulement sont dans ce cas; les autres sont, soit des fondations privées non commerciales, soit des institutions publiques ou semi-publiques.

L'activité des organismes européens de sondages est très honorable, comparativement à celle des meilleures organisations américaines, qui sont leurs aînées souvent. Deux organisations seulement font moins de six enquêtes d'opinion par an. Six en font un nombre compris entre 6 et 10, trois en font de 10 à 20, et quatre en font plus de 20 annuellement. Ceci naturellement n'inclut pas les enquêtes commerciales, ou les études spécialement commandées par des organismes ou des personnes intéressées, et qui ne donneront pas matière à publication.

Dans la plupart des cas, les résultats des enquêtes sont communiqués à la presse. Quatre instituts au moins disposent de contrats de presse. Mais, en outre, neuf instituts ont leurs organes propres. Deux publient plus de douze bulletins par an; deux ont une publication mensuelle; les cinq derniers publient moins fréquemment.

Le nombre des sujets interrogés au cours d'une enquête varie, suivant les instituts et suivant les enquêtes, entre 1.500 et 5.000, avec une moyenne d'environ 2.500. Presque tous les instituts utilisent la méthode de l'échantillonnage stratifié, sauf un, qui utilise l'échantillonnage au hasard. Tous les instituts qui pratiquent l'échantillonnage stratifié choisissent toujours leurs sujets en tenant compte du sexe, de l'âge, de l'habitat et de la région. Mais on peut observer une tendance, parmi les instituts les plus récents, à s'écarter du principe de choix par niveau de fortune, toujours assez incertain, et peut-être de moins en moins significatif en Europe, et à le remplacer par un choix par professions. C'est actuellement le cas d'au moins six instituts sur treize (14 pratiquent l'échantillonnage stratifié; les renseignements manquent à cet égard sur l'institut hongrois).

Ajoutons enfin que l'équipement matériel et en techniciens des organismes européens de sondage est en général bon ou très bon. Presque tous font usage des facilités du dépouillement mécanographique. Deux instituts disposent de moins de 100 enquêteurs. Trois en ont de 100 à 200, quatre de 200 à 300, et six en ont 300 ou plus.

On peut donc dire que l'Europe commence à disposer, depuis la fin de la guerre, d'un ensemble de stations pour l'étude de l'opinion, et que cet ensemble présente une bonne valeur technique, suffisamment homogène déjà. J'ai intentionnellement omis de mentionner les travaux d'une organisation britannique privée déjà encienne, qui ne fait pas à proprement parler des sondages, mais dont les enquêtes qualitatives d'opinion sont fort intéressantes; je n'ai pas parlé non plus de la section de sondages qui existe au sein de l'Institut National de la Statistique et des Etudes économiques, du gouvernement français. Cette derniere organisation, qui évite d'aborder les questions subjectives, ne rentre donc pas dans le cadre de cet exposé; elle ne manquera pas d'apporter cependant, sur le plan technique, sa contribution à de nombreux problèmes que pose quotidiennement la pratique des sondages.

D'ores et déjà, le fonctionnement et les recherches particulières des instituts européens d'opinion ont amené ou promettent d'amener à brève échéance des progrès de méthode intéressants. Ces progrès sont d'ordre très divers et d'importance très variable. Certains peuvent paraître négligeables et servir cependant de point de départ à des développements considérables. D'autres sont peut-être destinés à amener des déceptions. Je ne prétends pas à l'infaillibilité dans le choix que je vous en présenterai ici. D'un autre côté, je ne me flatte pas davantage de ne pas commettre d'erreurs ni d'omissions, et je m'excuse à l'avance pour cette possibilité auprès de tous mes collègues présents et absents. J'insisterai trop probablement sur les travaux français; ce sont, par la force des choses, ceux que je connais le mieux.

Au crédit de l'Institut britannique d'opinion publique, nous devons certainement porter d'ingénieux perfectionnements pratiques, notamment dans la rationalisation de la correspondance avec les enquêteurs et de la technique mécanographique. L'Institut britannique, doyen des instituts européens, possède une expérience consommée dans la pratique des sondages, et a acquis, dans les domaines où il opère, la maîtrise presque absolue de son instrument. Il a servi de modèle à de nombreux instituts nouveaux, notamment les instituts nordiques. Parmi plusieurs autres recherches, l'Institut néerlandais se signale à notre attention particulière par l'étude qu'il poursuit depuis un certain temps, des facteurs psycho-morphologiques des opinions. Il appartenait aux compatriotes de Heymans et de Wiersma de penser à chercher dans la caractériologie individuelle la source de la relative incertitude à prévoir les opinions d'un individu socialement déterminé. S'il en est bien ainsi, les recoupements traditionnels, tous de nature sociologique, de l'échantillonnage stratifié, devront à l'avenir s'accompagner de consignes aux enquêteurs, relatives au choix des sujets d'après leurs apparences physiques. Ce sont là des perspectives pleines de promesses.

Le jeune organisme de sondages de Milan, Doxa, est particulièrement actif en ce moment sur les problèmes pratiques de l'échantillonnage. Il a conçu d'une part une méthode d'échantillonnage stratifié qui serait fondé sur la localisation des différentes classes d'individus selon les heures de la journée. D'autre part, il mène des expériences de tirage au sort pour échantillonnages topographiques, à comparer aux résultats de l'échantillonnage stratifié. Peut-être contribuera-t-il à apporter, du côté expérimental, quelque lumière à l'irritante querelle des partisans de l'une et de l'autre méthodes.

L'Institut Français s'est lancé, lui aussi, à l'attaque de plusieurs problèmes pratiques. Je me limiterai à quatre questions, deux psychologiques, et deux statistiques. En premier lieu, il étudie depuis plus d'un an une méthode de tests psychologiques pour la sélection des enquêteurs; en second lieu, il a conçu, et finit de mettre au point, une méthode d'apprentissage des enquêteurs par disques de phonographe. Du côté de la pratique statistique, l'Institut Français a réalisé une table de pourcentages commode qui accélère considérablement les calculs arithmétiques dans de nombreux cas; et il a réalisé des abaques qui rendent très rapide le calcul de 2 comme test de signification des différences et des associations. Il a à l'étude un appareil qui sera destiné à montrer immédiatement le caractère significatif ou non significatif des différences et des associations.

Parmi les développments récents de l'étude des opinions en Europe, je ne dois pas manquer de signaler, en terminant, les efforts persévérants des instituts européens pour coordonner leurs recherches et l'étude internationale des opinions. Les accords bi-latéraux nombreux conclus antérieurement à 1947, se sont intégrés au congrès de Paris, en janvier 1947, par la création d'une commission européenne de l'opinion publique, qui groupe les instituts de l'Europe occidentale. Le congrès de Loxwood (Angleterre), en mai 1947, a de plus abouti à la création d'une association internationale des instituts d'opinion publique, qui lie huit instituts d'Europe à quatre instituts des Amériques et d'Australie. Puisse l'Association internationale des recherches sur l'opinion publique, préparée au début de ce mois de septembre 1947 au congrès de Williamstown, resserrer encore les liens de travail qui unissent dèjà tant de chercheurs et d'institutions en Europe et dans le monde, non seulement pour le bien de la science psycho-sociale, mais encore pour ce qui est le but le plus important de tous nos efforts, une compréhension réciproque enfin adéquate des difficultés et des aspirations des peuples par leurs dirigeants et par eux-mêmes.

Résumé

The study of opinions has not a very long history. One has only to go back to the decade from 1920-1930 to the first experiments which took place, both in the universities and in the field of market studies. The appearance of the first scientific investigation of public opinion

does not go back much before 1935. All of these attempts first appeared in the United States. On the other hand, if we confine ourselves to the year 1947 and if we prepare a geographic map showing the countries which are making studies of public opinions, we see that the maximum density of research is in Europe. There has been recently, therefore, a considerable growth in the study of opinions in Europe.

There exist 15 opinion institutes in Europe, 11 of which have been founded since 1944. These 15 institutes are located in twelve countries: Germany (2), Belgium, Denmark, Finland, France (2), Great Britain, Hungary, Italy, Norway, The Netherlands (2), Sweden, and Czechoslovakia. A second institute exists in Norway but the date of its founding is not known and it is not included here. It appears that there is also a second institute in Belgium and one in Switzerland; but these two apparently have not developed beyond the experimental stage.

Nine institutes are commercial or privately-owned organizations, 4 publicly administered organizations and 2 non-profit organizations. The training of the director is as follows: economics 5, business and public relations 3, statistics 2, psychology 2, sociology 1, other 2.

The activities may be shown by the following figures:

Number of Surveys Per Year	*Number of Institutions*
Less than 6	2
From 6 to 10	6
From 10 to 20	3
More than 20	4
Publications	
More frequently than monthly	2
Monthly	2
Less frequently than monthly	5
Without own publication	6
Number of Interviewers	
Less than 100	2
From 100 to 199	3
From 200 to 299	4
300 and over	6

All of the institutes use the quota method of sampling, with the exception of one which uses area sampling. Stratification is always made by sex, age, type of dwelling, and geographic region. Moreover, seven institutes use as a principle of sampling the level of income; five, occupational status; one institute uses both criteria. Data are lacking on one institute.

Allowing for errors and omissions resulting from the lack of complete knowledge by the author, the principal progress realized from the research undertaken in the field of opinion studies by the European institutes to date in 1947 is as follows:

British Institute of Public Opinion (BIPO): correspondence with the interviewers; mechanical sorting.

Netherlands Institute of Public Opinion: morpho-psychological factors of the opinions.

Italian Institute of Public Opinion (DOXA): recent developments in stratified sampling; experiments in area sampling.

French Institute of Public Opinion (IFOP); research on the preliminary selection of interviewers by psychological tests; training of interviewers by phonographic records; practical problems of statistics: simplified table of percentages, abacuses to speed up the calculation of chi-square, study of an apparatus designed to show immediately the significance of differences and of associations.

Finally, it is important to note the persevering efforts of the European institutes to coordinate their theoretical research and the international study of opinions. The numerous bi-lateral agreements concluded prior to 1947 were integrated at the Paris Congress in January 1947, by the creation of a European commission of public opinion which groups all of the institutes of continental Europe. The Loxwood Congress (England) in May 1947, resulted in the creation of an international association of institutes of public opinion which unites eight institutes of Europe with four institutes of the Americas and of Australia.

PUBLIC OPINION RESEARCH IN LATIN AMERICA

by William Herrington (*United States*), *President*
and
Alberto Castelli (*Argentina*), *Vice President*
International Public Opinion Research, Inc.

PRESENT STATUS OF PUBLIC OPINION RESEARCH IN LATIN AMERICA

Public Opinion research in Latin America is in a very initial stage, but this new science is advancing there more rapidly than in many other areas of the world. In recent years polling organizations have been established in Brazil, Cuba, Mexico, and Argentina, and opinion polls have been conducted in these countries and in Colombia and Venezuela. The work of most of these organizations has been of a preliminary or experimental nature, and has met with varying degrees of success. In most cases the polls have been restricted to segments of the population in the principal cities or in the higher socio-economic classes. Practical difficulties have prevented the taking of polls representative of the entire population of any country. For example, rural population has seldom been included, and the lower economic classes have not generally obtained proper representation.

FACTORS FAVORING THE DEVELOPMENT OF PUBLIC OPINION RESEARCH IN LATIN AMERICA

Several factors are at present working in favor of the development of public opinion research in Latin America. A growing interest in opinion research, and in the findings of opinion polls, has been shown in educational, business, and publishing circles.

In the universities, professors and students of statistics, economics, sociology, psychology, and related subjects are eager to be made aware of each new step in the growth of this as yet infant offspring of more mature sciences.

Alert businessmen including some who have heretofore made little use of statistics, have indicated a desire to be informed about the application of public opinion research to business problems. Business leaders are particularly interested in the use of public opinion research in public relations, where the attitude of the public toward an industry, a company, or its products, its distributors and services, may have a serious bearing on the ultimate, if not the immediate, success of an enterprise.

Editors and publishers, who are familiar with the interest of readers in countries where the results of opinion polls are published regularly as syndicated features in newspapers and magazines, are anxious to have such polls established in their own countries in order that they may use them as a means of increasing circulation. Some newspaper publishers in Latin America have attempted, within the limited knowledge and means at their disposal, to conduct public opinion research, but in general the results have not been very satisfactory when judged by strict standards of reliability and significance.

In addition to growing interest among influential groups, another factor working in favor of the development of facilities for public opinion polls is the demand of North-American and European business firms for better facilities for market research in Latin America. The fact should be mentioned that in the United States, Canada, England, France, India, Australia, and other countries where public opinion polls are now regularly made, this development followed the establishment of facilities for market surveys in behalf of business interests. As is well known, in the United States market surveys were conducted for more than a score of years before public opinion polling began in 1935. The concept "markets are people"* which required the development of means of finding out what people want, led naturally into the development of means of finding out what people think.

For many years business interests have been in need of more information about the consumer and industrial markets of Latin America. This need probably has been felt more outside of Latin America, by companies accustomed to having such information supplied to them by market research within their own countries, than within Latin America, where market surveys, until recently, have not been common.

This demand for more market information, derived from consumers and distributors themselves, has led numerous North-American and European companies to sponsor surveys and to employ and train the necessary national personnel for accomplishing them, at least in the most potential marketing areas. Such efforts serve to develop an appreciation among Latin-American business leaders of the important part which reliable market facts now play in the operations of most successful business enterprises. They serve also to prepare the way unintentionally but nonetheless definitely for public opinion research.

A third factor working in favor of the development of opinion research in Latin America is, of course, the present trend toward international opinion polls in which many areas of the world are to be represented.

This trend was discussed and furthered at the very recent Second International Conference on Public Opinion Research, meeting at Williams College in Massachusetts. At this conference a preliminary organization was set up to make preparations for a World Congress for Public Opinion Research to be convened in 1948. Thirteen nations were represented at the recent meeting, and many more are expected to be represented at the projected meeting next year. Elected president of the preliminary organization at this year's gathering was Jean Stoetzel, head of the French Institute of Public Opinion. Mr. Stoetzel is a recognized authority among opinion research experts, and the one who is best acquainted with the present status of the new science as applied in Europe. Among the vice presidents who will assist him in preparatory plans are: Dr. Laszlo Radvanyi, Professor at the National University of Mexico; Theodore Lentz, Professor at Washington University, St. Louis; David Glass and Mark Abrams of England; and Dr. Julian Woodward of the Elmo Roper organization, New York. Many others representing additional countries will participate in the work of preliminary committees yet to be formed.

This development of international polls is now largely the concern of individuals engaged in public opinion research, and it will continue for many years to be chiefly their concern, because of the extremely difficult if not insuperable problems which are encountered in any attempt to gather a consensus of public opinion on any question on a world-wide scale. However, the development of international polls arises also in response to a genuine need for them. Their use by agencies of the United Nations may be a somewhat distant possibility, especially in view of the fact that such polls are not now among the instrumentalities of national governments, but such use can be conceived of as a desirable possibility. Some preliminary trials have been made already by the Gallup affiliates in attempting to convey the same questions to people of several nationalities, and experimental work in this direction will be accelerated by the world organization for opinion research. When the time for international polls of public opinion arrives, the reliably recorded opinions of Latin-American peoples will be necessary among the opinion of other peoples of the world. This trend toward world opinion polls is an encouragement to the growth of public opinion research in Latin America today.

Factors Delaying Opinion Research Development in Latin America

Those of us especially who are engaged in the statistical and social sciences may think it worth while to consider, on the other hand, some of the factors which appear to be delaying the development of this modern science of opinion polling in Latin America.

Perhaps no attempt should be made to mention these delaying factors in any order of importance; however, the need for more teachers and students of the technical aspects of public opinion polling does appear to be among the primary blocks to progress. This need for experts d students acquainted with the principles, current practices, and remaining problems of this science is by no means limited to Latin America; it is world-wide. A significant peculiarity of this need in Latin America is the fact that the majority of the problems in practice have to be resolved at the local level within the countries in which the polls are made. These problems must be solved by individuals who are not only well-grounded in the science of opinion research but who are, also, most familiar with the available national and local means of solving them.

Fundamental in the knowledge of such experts in Latin American countries must be a thorough understanding of sampling theory and practice, questionnaire design and pre-testing, interviewer training and supervision, advanced coding and tabulating processes, significant means of analysis and reporting. Provision must be made in questionnaires and analyses for showing the effects of knowledge or lack of knowledge on opinion, the reasons for opinions, and the degree of convictions with which they are held.

Unfortunately, in the short period of its existence, public opinion research has become a very complicated science without adequate textbooks from which it can be learned. Many helpful studies have been published, especially in technical journals, but every conference of polling experts yields additional evidence of the fact that teachers and practitioners of the new science must indeed exercise great intellectual curiosity and considerable initiative if they are to keep themselves well-informed concerning the latest additions to the knowledge of and the most accepted practices in opinion research. Mr. George Gallup, a widely recognized authority in this subject, has stated that the chief factor delaying opinion research everywhere is the lack of qualified personnel, and this is true for America as well as for Europe and Asia.

The factor next in importance as a cause of delay in the development of public opinion research in Latin America is the lack of statistical data suitable for use in sampling, both on the national and the community or local levels.

As this group well knows, population and business census activities in many Latin-American countries have been handicapped in the past by a variety of irregularities with respect to timing, standards, scope, supervision, analysis, and publication, and particularly in the availability of funds. Such handicaps are by no means confined to Latin America—they plague the work of statisticians in all parts of the world and constitute a bond of sympathy. In our insatiable desire for facts and figures our profession is misunderstood, especially by politicians.

Public opinion research needs and makes effective use of the best existing census-type data. It calls for special organization of such data as well as for new kinds. In Latin America, as everywhere, the greatest need is for data which serve as criteria for determining the social and/or economic well-being of the population. Individual or family income statistics and home rental or valuation statistics may be mentioned as examples, but other criteria peculiarly indicative of socio-economic strata of population in Latin-American communities are needed as much or more. For example, data with regard to the kinds of bread eaten, the kinds of beds slept on, the kinds of roofs or floors of dwellings and the public services or conveniences used have been shown to be valuable tools to the sampling technician, particularly when these data are available at the community level. In the application of the newly devised area-sampling techniques, more community maps and listings of population are especially needed.

All opinion research workers in Latin America should work for the successful culmination of the plans for a simultaneous and standardized census of all Latin American peoples in 1950, the child of great promise long nurtured by the Inter American Statistical Institute.

A third factor which is likely to delay the development of public opinion research in Latin America is the tendency to regard this new branch of statistical activity as primarily a tool of business rather than as an aid to the social sciences. Social scientists and statisticians of all kinds in Latin America should recognize the place of public opinion polls in their work; should welcome this means of furthering their studies and attaining their objectives. If opinion research is left dependent for its support on business and commercial interests in Latin America, there will be a considerable delay. Those interested in this development should seek the support of public and private institutions. Such support will enable carefully designed experiments to be carried on and will bring Latin America to the forefront among the present users of this means of measurement.

FUTURE OF PUBLIC OPINION RESEARCH IN LATIN AMERICA

What, then, may be said of the future of public opinion research in Latin America? In considering this prospect, the authors of this brief discourse would like first to pay a tribute, and second to offer a warning.

Among the statistical and social scientists of Latin America today there is a great eagerness to learn more about and to make more effective use of public opinion research. The importance of this development is realized by many teachers and students in the universities and by directors and officials of government or private agencies.

Requests are constantly being received for more books and literature on the subject, and such material is being widely circulated and thoroughly studied. Many Latin American authorities and students have used the more direct means of learning by personal association with those who are teaching or engaging in opinion research. The teaching of sampling methods has begun in the universities, and technical treatises by Latin American authors are appearing in statistical and social science journals.

These are indeed encouraging signs, and the delegates of Latin America to this conference are to be congratulated on the alacrity with which they and their colleagues have laid hold of this new opportunity to advance the profession through the use of statistical science in public opinion research.

The advice of caution springs from several sad chapters in the experience of opinion researchers in the United States and other countries where polling has become an accepted practice. Public opinion research wherever it is practiced by authorities today is an exacting science. In all of its five or six methodological divisions the complexity is increasing rather than diminishing. It is true that public opinion polls in some countries have attained a high degree of reliability, but it is also true that the precise practices which have proved reliable in one country cannot be merely transferred to another.

Continuing on the warning note, the false starts and blunders which have accompanied the development of public opinion polls elsewhere can be avoided in Latin America if workers in this science will undertake the experimental work necessary to determine how the principles and most advanced practices can be interpreted for use under widely various conditions.

Those who engage in public opinion polling in Latin America, also, like professional opinion researchers everywhere, must guard against fly-by-night opinion research organizations and charlatons of every stripe, using unsound methods for selfish or even dishonorable purposes without regard for the reputations of those who endeavor to serve the common good and to maintain the highest standards. To prevent such retarding and damaging influences, the continuation of widespread educational efforts—both public and professional—is vital; and a special training program in opinion research, under the guidance of the Inter American Statistical Institute is greatly to be desired.

Résumé

Le sondage de l'opinion publique en Amérique latine se trouve encore dans la phase initiale, mais fait de plus rapides progrès que dans plusieurs autres parties du monde.

Plusieurs facteurs exercent à présent leur influence en faveur du développement du sondage de l'opinion publique en Amerique latine. Parmi les facteurs les plus influents il faut mentionner un intérêt croissant dans le sondage de l'opinion publique et dans les résultats des sondages, manifesté par les milieux de l'enseignement, des affaires et de la presse ainsi que des demandes toujours plus nombreuses, de la part de firmes commerciales de l'Amérique du Nord et de l'Europe, d'obtenir de plus grandes facilités pour l'analyse des marchés en Amérique latine.

Un obstacle dans la voie du progrès est le manque d'instituteurs et d'étudiants des aspects techniques du sondage de l'opinion publique en Amérique Latine, car la majorité des problèmes doivent en pratique être résolus dans le pays où le sondage a lieu. Un autre obstacle sérieux est le manque de données statistiques, tant au niveau national que local, pouvant servir de base à l'échantillonnage.

On peut aussi considérer comme une cause retardatrice, la tendance de considérer cette nouvelle branche d'activité statistique comme un instrument d'hommes de commerce et d'industrie plutôt que comme une aide pour les sciences sociales.

Cependant, il y a des signes d'avancement constant. Dans le domaine de la statistique et des sciences sociales on observe un grand intérêt chez les savants de mieux connaître et d'utiliser d'une manière plus efficace le sondage de l'opinion publique. Des demandes de recevoir plus de publications sur la matière parviennent constamment et nombre d'autorités et d'étudiants en Amérique Latine ont recours aux moyens plus directs d'étude en s'associant avec ceux qui enseignent le sondage de l'opinion publique ou se livrent à des recherches sur cette matière. L'enseignement de méthodes d'échantillonage a commencé dans les Universités, des traités techniques sont publiés par des auteurs américains latins dans des périodiques scientifiques.

La prudence est recommandée à ceux qui se livrent à des sondages de l'opinion publique en Amérique Latine, prudence que motive l'expérience acquise dans les Etats-Unis et autres pays ou le sondage de l'opinion publique est plus largement pratiqué. Bien qu'il soit exact que, dans certains pays, les sondages de l'opinion publique ont obtenu un degré élevé d'exactitude, il est également vrai que les techniques élaborées dans un pays, ne peuvent pas être utilisées dans un autre pays sans adaptations et modifications prouvées nécessaires par la recherche expérimentale.

Pour favoriser le développement des travaux de sondage de l'opinion publique en Amérique Latine, une action entreprise avec l'aide de l'Institut Inter-Américain de Statistique serait désirable.

UNITED STATES METHODOLOGICAL RESEARCH IN PUBLIC OPINION

by Samuel A. Stouffer

Laboratory of Social Relations, Harvard University (United States)

Research to improve the methods of public opinion analysis is considered by American scholars and practitioners as urgently important, for two reasons. First, the public opinion poll is now beginning to be recognized for what it really is—an instrument invaluable to the preservation of democracy. Second, public opinion studies properly planned and analyzed can help lay the basis for a science of social psychology. Before considering explicitly some of the recent American developments in public opinion research, let us look briefly at these two powerful incentives.

Democracy, to survive in a complicated industrial age, requires, among other things, swift and efficient methods of communication *upward* from the people to their leaders as well as *downward* from the leaders to the people. In both government and industry, the weakest part of the communication process has been the lines upward from the people to the leaders. Even such autocratic organizations as the Army recognized the need in wartime for better channels by which the desires and criticisms of the man at the bottom could be conveyed to top management. If this were true in the Army, how much greater is the need of democratic government to hear reliably from the people and for industrial management to hear reliably from workers or consumers.

The traditional methods by which the people communicated to their leaders—through the press and radio, through word of mouth and letters, through representation of organized pressure groups, and, of course, through free elections—will never be outmoded as long as democracy lives. But these methods have now been supplemented by a new method, swifter and potentially more accurate than any of the others—the public opinion poll. Here the public temperature can be taken on a calibrated instrument. What the clinical thermometer—whose universal use is less than a century old—has meant to the development of modern medicine, the public opinion poll can eventually mean to the development of efficient democracy in an industrial age. Newspapers and radio may not always mirror public sentiment—the majority of the press, for example, were against Roosevelt, the majority of the people for Roosevelt. Letters are notoriously unreliable, even when spontaneous and not stimulated by organized pressure groups.

Just before we entered the last war a count made of congressmen's mail showed that it was overwhelmingly against drafting men for national defense; the polls showed more than two-thirds of the public in favor of the draft. Organized pressure groups are a necessity to make sure that minorities will be effectively heard, but the fantastic claims of general public support often made by these groups needs to be exposed to the cold white light of the facts. Free elections are at the core of democracy, but elections involve parties, personalities, and highly oversimplified and emotionally charged slogans and seldom constitute an expression of knowledge or attitude about specific issues.

There are still some critics of public opinion polls who deplore them precisely because they are so effective in revealing the public mind. We can ignore those whose objections are based on embarrassment when the facts puncture the balloons of exaggerated or misrepresented claims made in support of pet causes. Harder to ignore are those sincere critics who have a low opinion of the public mind and public tastes and who fear that frequent and accurate temperature readings will encourage catering to the lowest common denominator and discourage action which, though not necessarily popular at the moment, is ultimately for the public good. That this is happening both in government and in industry, could not easily be refuted. But it can also be pointed out that efforts to educate the public are likely to be much more effective if based on accurate knowledge of the public's specific misconceptions or alleged low tastes than if based on no information better than a hunch. And, here, lies one of the greatest opportunities of public opinion polls—not so much to show whether the public favors or opposes a specific issue as to show how well informed the public is about the issue and how specific kinds of misinformation are related to specific opinions.

But the immediete practical job of making democracy more efficient, both politically and industrially, by improving the communication upwards from the masses to the leaders is only one of the incentives for improving our techniques. Another is the need of developing social psychology into a science.

Of course, there are those who think that social psychology already is far advanced on the road to being a science. Avoiding quarrels over definitions of the word science, we probably can get considerable agreement, however, on the proposition that social psychology has few general theories which have been verified with such rigor that they can be applied with confidence to concrete situations. Some of the new tools developed in public opinion studies promise to speed up the day when the engineering applications of social psychology will be applications of tested theory and not simply applied common sense wisdom.

American research to improve the techniques of public opinion studies has by no means developed in isolation. It has drawn on work done by scholars throughout the world. We shall consider three areas of research: (1) the sampling of populations, (2) the sampling of items, and (3) analysis of opinion dynamics. The first of these, in particular, illustrates the great debt which research in the New World owes to scholars from the Old World.

Public opinion research is made possible because a relatively small sample properly drawn will yield cheaply and quickly information approximating what would be obtained from a survey of an entire population. Sampling theory rests on the solid base of decades of fundamental mathematical work by scholars in France, Germany, Italy, Scandinavia, Great Britain, and other lands. In particular, the developments of the past quarter of a century, with their focal center in England, have laid down the essential lines along which current statistical thinking travels. The largest American contribution has been in what might be called the engineering aspects, though, as the pages of such journals as the *Annals of Mathematical Statistics* eloquently testify, American scholars have also widened and deepened the fundamental thinking.

Most polls in America today use what is called the quota method of sampling. The sampling is stratified by selecting demographic characteristics, such as region, age, sex and economic status, proportional to known population figures. The interviewer in a given locality is assigned a quota—so many in each age, sex, and economic group, for example. From this point on the interviewer is on his own. He makes his own selection of respondents within this quota. Of course, he is warned about various ways in which bias can creep in, and there is often a retrospective attitude check on his work. For example, if he asks, "Whom did you vote for in the last presidential election?" the figures from a national sample if unbiased, should closely approximate the known national figures.

The remarkable success which the polls have had in forecasting elections is evidence that for many practical purposes the quota method of sampling may be good enough. However, there is at least one glaring weakness in the quota method. Since the final choice of respondents is discretionary with the interviewer, it cannot be said that it is at random. The magnificent paraphernalia of mathematical statistics do not strictly apply in the quota sampling operation.. When predictions can be checked, as at elections times, a practical test is available, but ordinarily there is no way of specifying precisely the sampling error or applying theories of statistical inference. Moreover, there are varieties of situations where biases can enter in.

Some agencies, notably in government, have discarded the quota method in favor of what has been called the area method of sampling. The choice of addresses of respondents is no longer left to the discretion of the interviewer, who is instructed to call at a specific list of addresses drawn at random by the central office from a stratum of a prefabricated sample. There are special problems, such as the control of what respondent in a household a caller interviews or such as complications due to recall and substitutions, but the advantages of the area method lie primarily in the essential randomness of the ultimate selection of respondents.

A study financed by the Rockefeller Foundation and sponsored by a committee of practitioners and academicians appointed by the Social Science Research Council and National Research Council is attempting to assess the relative advantages and disadvantages of these two methods of sampling and of combinations of the two methods. This study, now nearing completion under the direction of Professor Frederick F. Stephan of Princeton University, should clarify many of the issues, including the exceedingly important one of comparative costs, although many more years of research and experience will be necessary before definitive answers to all the complicated questions may be expected. The same committee of the Social Science Research Council and National Research Council is also sponsoring an extensive study of interviewers' bias, which is under the direction of the National Opinion Research Center, using funds supplied by the Rockefeller Foundation and Mr. Marshall Field, who has had a long interest in the improvement of public opinion studies. This project involves, incidentally, a good deal of actual controlled experimentation in which efforts are made to induce biases in order to see more clearly how they operate.

Other studies are in the planning stage, notably a study of the advantages and disadvantages of repeated interviews through the use of panels. An idea of the large scope of practical work in this field—such as consumer inquiries for industrial concerns including publishing houses—now going on in the United States may be had by considering the fact that upwards of a hundred panels are operated currently by New York market research organizations alone. Through panels it is possible to study changes in attitude of the same person and thus relate changes directly to other characteristics of the individual. The study now being planned of panel operations may open some new vistas not only for practitioners but also for academicians interested mainly in the developments of social science.

Having discussed some of the directions of interest in research which is concerned primarily with the theoretical and practical problems of the sampling of populations, let us now turn to another area, namely, the sampling of items. There are an indefinitely large number of ways in which questions can be asked to elicit a given experience or attitude. Moreover, it is easy to multiply examples of how seemingly similar questions will yield strikingly different responses. The better commercial practitioners are very well aware of this and when possible pretest their questions before going into the field, as well as use duplicate or triplicate forms of question wording. Nevertheless, there is often a wide latitude for interpretation, not only of the meaning of a question but also of the intensity of the response.

To overcome these hazards, there have been two main lines of development, quite different in nature. One approach, which perhaps has reached its highest point of development in the work of Dr. Rensis Likert of the University of Michigan, and his associates, is to introduce much more informality into the interview than in conventional polls. The interviewer, more highly trained than the conventional interviewer, does not ask a fixed set of questions, but in effect, visits with the respondent, asking informally whatever questions seem appropriate to the situation and finally building up from the respondent's detailed responses an impression of what his opinion on a subject is and how strongly it is held. A very detailed report of the conversation is sent to the central office where it is analyzed and summarized by expert coders. This operation, which can cost up to $25 or more per interview, is thought too expensive by many practitioners for wide-scale use, but it has yielded some very convincing results and seems to be of particular advantage in situations where attitudes are not sharply focussed or where numerous contingencies must be separately evaluated before a prediction can be safely made. An example of the type of situation in which the open-ended interview has been notably successful is in the investigation of respondent's general buying expectations in the year ahead. Although predictions made by this procedure can be checked against subsequent fact and although on numerous occasions incidental information obtained by informal interviews has been found to correspond closely with independent data (for example, money said to be in the bank), the great amount of discretion left to the interviewers and coders has opened some perils of its own, as compared with the more conventional methods. That the method is ideally suited for preliminary scouting studies on small samples or for supplementary spot checks to help interpret material much more cheaply obtained by the conventional methods, it is generally agreed.

A different approach to the problem of surmounting the hazards of individual check-list items is to use several such items in some form of a scale. There are many possible kinds of scale. One scale may merely order respondents. Another may order respondents on either side of a zero point. Another may order respondents on a scale with a metric like a thermometer on which the zero point is arbitrary but in which intervals are equivalent—for example, the interval between 15 and 10 could be psychologically equivalent in magnitude to the interval between 10 and 5. Finally, in the latter scale a psychologically meaningful rather than arbitrary zero point might be introduced.

A detailed review of the American research on scaling would carry us far beyond the scope of this paper. The commercial polls have as yet made little use of scales, but research on the subject is very much alive. The best known of earlier attempts were those initiated twenty years ago by Professor L. L. Thurstone of the University of Chicago, adapting psychophysical techniques first developed in Germany. Two main methods were developed—one based on paired comparison and one based on equal appearing intervals. Using with the paired comparison what Thurstone called the law of comparative judgment, one could obtain and test for interval consistency a scale in which intervals were made psychologically equal.

During the war, the need for a simple technique which would employ a small number of check-list items led to a reconsideration of basic theories underlying attitude measurement. In the Research Branch of the Information and Education Division of the War Department, which was polling American soldiers all over the world, a new approach was developed, mainly under the leadership of Professor Louis Guttman of Cornell University.

Guttman's model is quite simple. Suppose you ask a man three questions, each of which he can answer Yes or No:

"Are you over 6 feet tall?" Yes. No.
"Are you over 5 feet 6 inches tall?" Yes. No.
"Are you over 5 feet tall?" Yes. No.

If the answers to all three questions are consistent, respondents will be perfectly ordered into four groups, those 6 feet and over, those 5 feet 6 to 6 feet, those 5 feet to 5 feet 6, and those under five feet. Moreover, if we know a man's rank—that is, in which of the four groups he belongs —we can from the mere knowledge of the rank reproduce his answers to all three questions. Thus the man who is in the group 5 feet 6 inches to 6 feet must answer the first question No and the other two Yes.

This model can be used with attitude term items, if the attitude is on a single dimension. In fact, its greatest utility may turn out eventually to be the identification of the context in which a specific attitude item belongs. Response to a series of items must be such as to permit their ordering in a cumulative way. Thus, a person who checks Yes to Item A, to which the fewest people check Yes, must also check Yes to all other items in the list. Scales developed by this method were quite extensively used by the War Department, and have since been tried on other studies. It should be noted that the procedure yields only an ordering and does not yield a metric. An extension of the technique through the introduction of an intensity function, led to the development of a zero point. In instances when the exact zero point could not be reliably determined, a "zero region" could be indicated.

Also growing out of the War Department's work was a different approach to the scaling problem. This is the latent attribute technique developed by Professor Paul F. Lazarsfeld of Columbia University. Taking off from the earlier Spearman work in England on factor analysis, Lazarsfeld has developed, for what seems to be the first time, a sound technique for factor analysis with qualitative data. Applying his model to attitude data, he considers the response to an item belonging to a one-dimensional scale to be an expression of a latent attribute plus a residual. If the residuals are independent, it becomes possible to order respondents to several items according to the probability with which they possess the latent attribute. This approach has very interesting possibilities. In practice it has been found that items which form a scale by the Guttman techniques do not form a scale by the Lazarsfeld technique, and vice versa. A good deal of progress has been made in analyzing, mathematically, the relationship between these two different models and also between them and the earlier work of Thurstone. The whole subject will be reviewed in detail in Volume IV of a forthcoming series of studies growing out of the work of the Research Branch of the War Department. This report is prepared under the auspices of the Social Science Research Council, 230 Park Avenue, New York City.

It is possible that still better approaches to the scaling problem will be developed. World War I gave a considerable impetus to the development of intelligence measures and achievement measures and it is likely that World War II will have given a similar impetus to development of measurement of opinion and attitudes. A large amount of concerted work on this problem may be expected in the next few years.

We have spoken briefly of research involving the sampling of population and the sampling of items. We close now with a few words on studies of the dynamics of opinion. In spite of the fact that a number of textbooks and treatises are available on public opinion, and in spite of the wisdom of politicians and the insights of such experts as the late author of *Mein Kampf*, the tested hypotheses are few indeed.

There is now in America a spirited interest in reformulating the whole theory of the formation and changing of public opinion in operational concepts which can be put to empirical test. University centers in which there is particularly active concern with the problem include Columbia, Princeton, Michigan, Cornell, Chicago, University of Washington, Yale, and Harvard. The subject is being approached from several directions, of which four may be briefly noted. One is to assemble in a single archives the punch cards from hundreds of surveys—as has been done by Professor Hadley Cantril at Princeton—and by exploring this vast mine to obtain comparative data which will at least give us some preliminary answers to general questions. Another is to make careful repeated interviews on a large panel of respondents at a critical period—as was done by Lazarsfeld of Columbia in an election year—and note who changed his opinion and with what that change in opinion seemed to be related. Still another approach is under way at the Harvard Laboratory of Social Relations, where a handful of individuals has been studied with very intensive clinical techniques over a period of several months, involving spending thirty hours with each respondent. Psychologists, anthropologists, and sociologists have collaborated in this study, which is merely a preliminary one for the purpose of formulating more promising hypotheses, some of which are now being further explored with a larger sample. Finally, progress has been made in the testing of some hypotheses by actual controlled experiments. There is quite an extensive literature on experimental studies of attempts to change attitudes by movies, film and radio, or in the classroom. In the series of volumes reporting on the work of the Research Branch in the War Department during the war, Volume III will be devoted to problems of making experimental studies of the effects of film and radio in changing attitudes. Further experiments are under way, notably in the Department of Psychology at Yale, under the direction of Professor Carl Hovland and Mr. Arthur Lumsdaine. Social psychologists are coming more and more to realize that about the only sure test of a hypothesis in our field is provided by the controlled experiment. There are many difficulties particularly in designing situations which are not defeated by their artificiality. And this work is very expensive. But it must be done if we are to have not mere insights and guesses but a science of social psychology.

In all too cursory a manner we have reviewed some of the developments in public opinion research in America. In comparison with millions spent annually in application of the methods, the money spent in research to improve the methods is, of course, quite small. But there never has been such an impetus for research as at the present time. Much is owed to the contribution of scholars from other lands. By keeping open our international lines of academic communication, through journal publications, through UNESCO, through gatherings such as this today, we can increase our exchange of ideas and experience. And all of us can share in the satisfaction of knowing we are forging new tools of democracy which can help—if ever so little—both to build social science and to contribute, through strengthening democracy, to the peace of the world.

Résumé

La recherche en vue d'améliorer les méthodes de sondage de l'opinion publique est considérée aux Etats-Unis tant par les théoriciens que par les praticiens, comme étant d'une importance capitale et céla pour deux raisons. Tout d'abord l'on commence à reconnaître que le sondage de l'opinion publique est un précieux moyen de conservation de la démocratie, complément des méthodes traditionnelles par lesquelles le peuple exprime ses désirs et ses opinions à ses dirigéants. En second lieu, les études de sondage de l'opinion publique peuvent aider à poser les fondements d'une science de la psychologie sociale. L'auteur envisage trois champs d'investigation concernant la méthodologie de sondage de l'opinion: (1) *La methode d'echantillonnage de la population,* avec des considérations sur les avantages et les désavantages relatifs des méthodes dites de *quota-sampling* et de *area-sampling* dans le sondage de l'opinion; (2) *Le choix des questions d'enquête,* avec une discussion du progrès récent et vraiment frappânt du développement de modèles mathématiques pour l'évaluation des attitudes; (3) *L'analyse de la dynamique de l'opinion,* au moyen de l'étude comparée des données de sondages d'opinion fournies par les archives; au moyen de sondages répétés montrant quel est le genre de personnes qui changent d'opinion; et par des expériences contrôlées d'influencer les opinions au moyen de films, de programmes de T.S.F., d'enseignement scolaire. Les recherches américaines ont fait grand usage des précédents travaux des savants européens, et en particulier de ceux qui concernaient le développement de la théorie de l'échantillonnage. L'usage, très répandu aux Etats-Unis, tant par les institutions privées que par le gouvernement, des techniques de sondage de l'opinion a fourni des encouragements, des données de base, une expérience pratique, et une aide financière à l'effort concerté actuel qui tend à améliorer la recherche méthodologique. L'usage de la technique du sondage de l'opinion dans l'armée américaine, qui fera l'objet d'un rapport en quatre volumes que doit publier le Conseil de Recherche des Sciences Sociales, est un exemple de la façon dont

les besoins pratiques ont stimulé le développement technique. De même que la première guerre mondiale avait donné une forte impulsion au développement, en Amérique, de l'amélioration des moyens de mesurer l'intelligence et les prestations, il est probable que la deuxième guerre mondiale aura pareillement stimulé le développement des techniques d'évaluation de l'opinion et des attitudes. Parmi les centres universitaires, où les recherches en ce domaine sont le plus activement poussées, se trouvent ceux de Chicago, de Columbia, de Cornell, de Harvard, de Michigan, de Princeton, de l'Université de Washington, et de Yale.

Discussion

Mr. Abrams suggested that the definition of public opinion research be broadened so as to include more than public opinion institutes. He indicated three broad categories in which work is currently being done in England: (1) the Government, concerned with attitudes affected by government social actions; (2) semi-governmental bodies, notably the British Broadcasting Corporation in the fields of program planning and listener research; (3) ordinary commercial organizations which do three types of work: (a) research work for the Government, especially on political issues; (b) market research and surveys in areas of public controversy; and (c) research for philanthropic foundations.

Mr. Abrams pointed out that public opinion research originated in England in social surveys, rather than from election forecasting as in the United States. Early surveys were in the field of housing.

Mr. Abrams indicated that most of the British survey-directors were originally economists; no outstanding director is a statistician or psychologist; their training has been primarily in business. There has been very little publication of results because work for the government and private commerce must be secret.

The bulk of British surveys, said Mr. Abrams, are made on quota samples. Some work has been done with systematic sampling from lists.

Mr. Abrams stated that for Britain quota samples are not wholly reliable because sufficient statistics are not available for the establishment of valid quotas. With respect to the control of interviewer and question bias, he cited two devices in use in England: (1)"dummy informants" —in a list sample, each interviewer is assigned a few respondents who have been drilled in advance on their answers (2) repeated interviews by the supervisor. A third development has been in the direction of intensive interviews with long questionnaires. This movement toward the study of the origin and purposes of opinion is, in Mr. Abrams' opinion, a result of the orientation of opinion research around problems of social policy.

Mr. Radvanyi discussed the work of the "Scientific Institute of Mexican Public Opinion," which he described as better established than similar organizations in Latin America. Employing students from the National University of Mexico, it makes surveys on national and international economic, social, political and cultural problems. A second organization, the "Institute for Studies in Public Opinion and Attitudes," investigates methodological problems and maintains contact with experts all over the world.

Mr. Radvanyi announced that the Mexican Institute, recognizing the importance of international opinion research, is publishing the *International Journal of Opinion and Attitude Research* and is preparing an *International Directory of Opinion and Attitude Research.*

Mr. Radvanyi proposed for consideration, in connection with international pooling, the following statements: (1) Identical procedures cannot be established for all countries. (2) Major conditions determining methods are cultural traditions, economic and social development, political situations, and degree and content of education and information. (3) Language differences may force the use of different questions to achieve the same effect in different countries. (4) Answers to the same question may have different meanings in different countries. (5) Different attitudes toward polls exist in different parts of the world, sometimes facilitating, sometimes rendering difficult the work of surveying. (6) "No opinion" and refusal to answer may have different meanings in different social and cultural backgrounds. (7) In some countries, for various reasons, false answers are given instead of "no opinion" or "no answer." (8) "Quota controls" valid for one country may not always be transplanted to another. (9) Work is needed to make "area sampling" feasible where maps and other data generally utilized are not available. (10) Study is needed to determine appropriate interviewing techniques for different countries. (11) The main concern must be to obtain valid and commensurable results in all countries. (12) The newly-formed World Congress for Public Opinion Research must work with the International Statistical Institute on the organization of international opinion surveys.

Mr. O'Grady stressed the importance of research on the choice of topics for opinion studies, to determine the priority of areas of socially significant attitudes. An example is the increasing attention to measurements of the extent of ignorance and misconceptions.

Mr. O'Grady drew attention to a field of research related to the study of opinion dynamics: the measurement of what opinion-forming forces are influencing given situations. Public relations research and communications research contribute in this direction.

Mr. Likert commented that he believed it unfortunate that the narrow term "public opinion research" has been used for the important social science research tools developed in the field of "survey research." In addition to opinions, data are collected by means of surveys on: (1) The extent of information, misinformation, and ignorance; (2) the behavior of people; (3) the experience of people; (4) motives; (5) economic, demographic, and social facts; (6) the relation between attitudes, information, etc., and behavior. Mr. Likert noted the use of surveys in measuring situations before and after experiments.

Mr. Likert made some general observations on the work being done in the United States, urging that commercial organizations be encouraged to publish (or make available for research workers) their results and especially their methodological findings. He concluded with the requests that sampling techniques be used which permit of the measurement of error, and that design of experiments and training of interviewers be improved. Interviewers must be enabled to establish and maintain rapport with respondents; something of the nature of free-answer interviews must be used to clarify what the respondents mean.

Mr. Tagliacarne added a few words concerning the Acta Institute in Italy, a private organization whose director is one of the few statisticians to hold such a position. The Institute, employing 300 interviewers, usually uses long questionnaires; it has studied various economic, social and political topics. Furthermore, the Acta Institute publishes its data with full detailed analysis.

Mr. Herrington referred to his experience in dealing with language problems in questionnaire development, through the use of preliminary interviews where the subjects of investigation are discussed informally. Another experimental technique has involved word-association.

Mr. Lina Campos made a concluding remark on the need for international collaboration in this field.

SOCIAL STATISTICS SECTION

Wednesday, September 17, at 2:00 p.m.

Social Security and Health Statistics

CHAIRMAN :

Ph. J. Idenburg

PROBLEMS IN THE DEVELOPMENT OF COMPARABLE STATISTICS IN THE SOCIAL SECURITY PROGRAMS OF DIFFERENT COUNTRIES

by Wilbur J. Cohen
Assistant Director, Bureau of Research and Statistics
Social Security Administration (United States)

Although social insurance systems in the modern sense of the term are now over sixty years old, no *comprehensive* international statistical series have been authoritatively developed which would enable valid comparisons to be made of the relative status and progress of the social security programs of the different countries. Many factors have contributed to this situation among which are changing economic and political conditions and frequently changing social security laws. But of primary importance to statisticians and economists is the fact that the absence of essential series on national income, labor force, family income, and related items has restricted international comparisons on a relative basis to a limited number of countries on a limited number of aspects, and to a very limited number of times, usually when very special studies or investigations have been undertaken.

Absolute figures on such important matters as coverage, contributions, beneficiaries, amount of benefits, and on similar topics have been collected and published from time to time by the International Labor Office. These have served as a valuable basis upon which students in various countries have compiled *relative* figures such as: the ratio of coverage under the social security program to the total population of different countries; contributions as a percentage of payrolls or national income;

beneficiaries as a proportion of the number of persons experiencing the different risks; or benefits as a percentage of average wages.

While absolute figures are essential in themselves for internal evaluation of the effectiveness and scope of social security programs in each country, because of the differences in the size of the various countries, the character of their economy, and the widely differing scope and level of adequacy of the social security programs, a valid statistical series of relative values should be carefully developed. Any such series of items would still require qualification and critical analysis before being widely used, but they would supply a greatly needed resource in the studies now underway in all parts of the world tending toward making social security universal in coverage and protection and simplified in administration.

One reason why in recent years it has become more difficult—and at the same time more essential—to make international comparisons is the diversity and complexity in the types of social security programs in operation. Twenty years or so ago, social insurance systems tended more along the lines of the " classical " type developed by Bismarck in Germany or Lloyd George in Great Britain. Today, the newer systems and those recently revised tend more in the direction of combining a number of programs together in a comprehensive, unified system with cash benefits, and services (medical, employment, rehabilitation), and with financing and administrative arrangements which are not easily classified. The Beveridge Plan in Great Britain and the New Zealand Plan are illustrative of the newer types of social security plans which by their unity may make detailed statistical classifications more difficult in some respects than under the older type of system.

In developing more comprehensive and unified social security systems it has become apparent that more than just the social security system of a country must be taken into account in framing, modifying, or evaluating social security legislation. Special systems for particular groups of workers, public assistance programs and programs for veterans, their dependents and survivors—all have a " social " purpose and must be taken into account.

I venture to predict that as the standard of living of a country increases, its overall social security program—that is, all public and private programs—will become more diversified and complex just as the economy in general becomes more complex as it becomes industrialized. The task of making international comparisons will become more complex, therefore, and all the more important.

This diversity in the institutional framework of social security means that more than one type of statistical series must be developed to show the changing, diversified, complex pattern. This calls for more basic data and simultaneously for more sample surveys, estimates, and adventuring into unknown fields. Statisticians will have to enter where both fools and angels fear to tread.

There is only one way to begin to learn to walk and that is to take the first step. A number of persons in different countries have been working on the development of national statistical series for their own social security programs. No valid international comparisons can be made without the national series. From the experience developed by those countries with national series we may be able to develop not only valid international comparisons but at the same time more useful revised national series.

It is to be hoped, therefore, that the appropriate national and international social security agencies in conjunction with the appropriate international statistical institutes and the International Labor Office can work together closely to evolve a statistical series which can form a valid basis for international comparison.

Some progress has already been made in the field of health statistics which are of importance to health insurance comparisons. Progress has also been made in the field of unemployment statistics which are of importance to unemployment insurance comparisons. Progress in various countries in the development of other statistical series on wages, national income, and labor force indicate that a real possibility exists for making such comparisons—and also for utilizing social security data more effectively in making estimates on such basic matters as the total wage and salary payments in the nation, national income, labor force, and total population.

The progress recently made at meetings of the Statistical and Medical Commissions of the Inter-American Committee on Social Security with representatives of the Inter American Statistical Institute and the International Labor Office indicates that there is an important task for statisticians, economists, doctors and other professional personnel interested in making valid international comparisons. This task is to set guides so that the newly established social security systems may take advantage of the experience of the older systems; so that all countries may profit from the many years of accumulated experience and establish statistical series which will be of value in the future comparison, evaluation and improvement of their programs.

Innumerable problems and difficulties confront those who attempt to define the bases upon which a series of relative statistical values can be compiled for international comparison. Those of us who have been concerned with framing the legislative provisions of social security programs or administering social security programs know the value which many statistical series have at a crucial point in determining policy, but which frequently cannot be obtained or if obtainable are often so limited as to be of doubtful value. A complete listing of all our needs would fill an immense volume. I can only suggest that it might be best to begin modestly and obtain agreement first on some essential and comparatively easy items. Then building on this experience agreement could be obtained upon the more complex items.

Illustrative of what I mean by the comparatively easy items are those relating to population coverage. Beneficiary coverage ratios are more difficult and ratios relating to the adequacy of benefits are exceedingly difficult. In the field of financing there are several ratios which are not too difficult to obtain although a longer time series is necessary for the ratios to be useful for international comparison. I shall refer to each of these four types of comparison very briefly.

Population Coverage Ratios

Although the newer social security systems—and amendments to the older plans—are tending in the direction of providing universal coverage, some program in every country excludes some persons from coverage. Even when the universal risks such as old age, death, or ill health apply universally, the risk of unemployment relates only to persons in the labor force. Originally, social insurance developed as protection to wage-earners but in recent years the scope of protection has broadened so that coverage against the universal risk is being applied to the self-employed (small businessmen, professional persons, farmers, artisans, independent merchants). In order to measure the scope of coverage of the plans of various countries it is important to have a measure of the *relative* number of persons given *protection* against a specified risk. The number of persons who have protection against a particular risk not only depends on the industrial, occupational, or regional coverage of the insurance plan, but also on the liberality or restrictiveness of the eligibility (and related) provisions of the system in relation to employment and compliance. Since the eligibility conditions usually vary as between different risks it is necessary to establish a coverage ratio for each separate risk. It is also necessary in some systems to select more than one date to obtain a valid international comparison because eligibility conditions may be related to calendar or fiscal years which may affect the ratio markedly, particularly in periods of rapidly changing employment conditions.

The problem of preparing a count or an estimate of the number of insured persons will be more difficult in those countries with many systems. Particularly in countries like the United States where many persons may be insured under more than one public system, an estimate is necessary to eliminate duplication.

A population coverage ratio can be published regularly once or twice a year only if reasonably accurate population estimates for such dates are available. For unemployment insurance, (and for temporary disability insurance) a labor force estimate is required—possibly with the self-employed and agricultural labor separately totaled. For old age insurance the adult working population could be used. For permanent disability insurance, survivors insurance,[1] and health insurance the number of insured and their dependents could be computed as a proportion of the entire population.

It is obvious that any population coverage ratios will not only involve a frequent count or careful estimate of the number of primary persons insured but also the number of dependents of the insured, and the labor force and total population both with at least a minimum number of classifications for more refined comparison.

BENEFICIARY COVERAGE RATIOS

Another broad type of comparison to which study should be given is what I have called "beneficiary coverage ratios." Such ratios would be the proportion which each type of beneficiary constitutes of the total number of persons experiencing a given risk. For old age, the ratio would be the percent which aged beneficiaries comprise of the population age 60 or 65 years and over depending on the retirement age. For survivors insurance, the measure could be the percent which paternal orphans represent of all such orphans under a given age. Other measures can be devised for other classes of survivors such as widows or parents but the most important single measure is for children.

For disability, the problem is somewhat different and more complex. Most disability systems are developed around the earnings loss concept, that is, they involve loss of earnings due to cessation of employment as a member of the labor force. The basis of comparison for disability might well be, therefore, both the total number of persons incapacitated (separately for the temporarily disabled and the permanently disabled) and the number of persons of adult age who were members of the labor force until their disability. In the absence of such an estimate, coverage could be expressed in terms of the number of disability beneficiaries per 100,000 persons employed, or per 100,000 persons (or males) between 18 and 64.

---

[1] Another type of interesting relative international comparison would be the proportion in each country which the face value of the survivors insurance represents of the grand total of all survivors (life) insurance (public and private). This type of comparison illustrates the kind of series which should be developed to show the relationship between governmental provisions for social security and private provisions with the same general objective. For old age, a few countries might be able to develop a ratio of the number of persons insured who also are insured under supplementary systems for additional protection.

The base for beneficiary comparison in the field of unemployment insurance would be the total number of unemployed and for children's allowances, the total number of children. For health insurance no single measure is appropriate but it may be that some study and experimentation would yield a minimum number of ratios which would be easy to produce and at the same time meaningful. For instance, the proportion of days of hospitalization for insured persons to the total number of days of hospitalization for the entire population might be of interest.

Comparisons in the health field must be so much more detailed than those for cash benefits that I have deliberately omitted any detailed reference to them in this paper. They are important and complex enough to warrant special treatment. However, I should like to refer again to the pioneering work on health statistics of the Statistical and Medical Commissions of the Inter-American Committee on Social Security and to point out that their recommendations will be discussed at a full meeting of the Inter-American Conference on Social Security in Rio de Janiero this fall.

ADEQUACY OF BENEFIT PAYMENTS

The most important single comparison among social security systems is with respect to the amount of the benefit. Since the various laws utilize many different formulae for the computation of benefits (flat rates, wage classes, percentage of wages) with varying increments for dependents and years insured, it is exceedingly difficult to determine the level of benefits paid and even more difficult—if not impossible—to determine the adequacy of the benefits. These matters are exceedingly difficult to determine in any single country. An international comparison is even more hazardous.

Is an international comparison of benefit payments then impossible? I do not think so. I believe that we could produce some useful ratios that would be meaningful in evaluating the programs of the different countries. Since the general objective of social security benefits is to replace lost income, it is wholly appropriate that the payment be related to some type of income in the country—wages, for instance. The basis of comparison would then become the percent which the average benefit (for a certain class or classes of beneficiaries) represents of the average wage in the country, or, if that is not available, of the average wage in manufacturing, or the average wage of persons contributing to the insurance system. Such an approach has the merit of holding constant wage levels, price levels, standard of living, and other measures of the well-being of the nation, and keeping the comparison to the place of the benefit amount within the frame of reference of the national standard of living.

It does not, however, enable one to judge the effect of such a benefit since the relation of income to a minimum standard of living might require a benefit of 100 percent of wages in one country and only 35 percent in another. Here, as in many other respects on the comparisons mentioned in this paper, careful qualitative analysis and other supporting statistical data must be used to supplement any single statistical series for effective use.

Study might be given to whether there would be any value in a comparison of average benefits to a statutory minimum wage. Because of the character of some national minimum wage laws it may not be practical to develop any such comparison generally.

The most meaningful comparison would be the relation of average income in beneficiary families from all sources (social security benefits plus all other types of income) to the average family income in the nation as a whole, or among certain types of families. This is an exceedingly difficult task. It requires careful sample studies of beneficiaries' incomes as well as statistics on family incomes for the nation as a whole. Such comparisons might be attempted for a selected group of countries for predetermined dates.

The study of the adequacy of social security benefits is the most difficult of all of the areas in social security because of the subjective character of "adequacy." Yet, some objective tests can be established. The combined efforts of many professions are necessary to accomplish the result. Studies leading to some acceptable standards of family income adequacy may make it possible to relate average income in beneficiary families to such standards. Here is an important and controversial area of work which awaits exploration. It may not be possible to achieve a satisfactory comparison along these lines, but the recent development of economic and social studies as a result of the interest in post-war reconstruction and the work of the United Nations may make it possible to attempt something that may produce some variation which will have some value.

FINANCES OF SOCIAL SECURITY

One question that is always asked about a social security program is, how much does it cost? Many answers are usually given to this question, but not always a very meaningful answer. The amount of money set aside in a particular year for benefits, or the amount spent on benefits, will not answer the question intelligently by itself. Here, as in other comparisons, more than one type of comparison is essential if we are to develop a full answer to the question asked.

One set of ratios that should be developed would be contributions and benefits each as a percentage of national income for single years, and as an average for five or ten year periods. For unemployment insurance an agreed upon period of eight or ten years could be used. For long run risks like old age, a much longer period might be used.

An average cost over a number of years might be developed by relating benefits to the taxable payroll for each insurance program.

SUMMARY OF DATA NEEDED FOR VALID COMPARISONS

The comparisons mentioned in this paper are merely illustrative and suggestive of the tremendous possibilities and opportunities which exist for advancing the work of evaluating and improving social security and its contribution to the national and international welfare.

Innumerable other comparisons have been made for selected social security programs or could be made among those with certain common features. It is, for instance, possible to develop many interesting and valuable comparisons among systems that have the common feature of flat contribution and benefit provisions; or, on the other hand, which have contribution or benefit provisions measured as a percentage of wages. But the task which I have tried to deal with is one of attempting to develop measures which could be applied to all types of systems.

Administrative personnel and costs is a subject in which there is a great deal of interest—and misinformation. The tremendous diversity in the patterns of administration, the interlocking of various social security programs, and the relation between social security and other public or quasi-private institutions makes statistics on administrative personnel and cost difficult to obtain and when obtainable easily subject to misuse. I have not the time here to go into this important subject but I believe it is important and worthy of separate consideration.

Although I have touched upon only a very few comparisons I should like to summarize the types of data needed to make international comparisons in the field of social security. By relating such a listing to the data now available in various countries we will have a kind of balance sheet of gaps in our information requiring attention.

First, with respect to the social security system itself, we would need data on the number of beneficiaries, by type of risk, classified by age, sex, familial status; the number of families in which beneficiaries live, by size and composition; the number of persons insured against specified risks, by risk, and by age, sex, and familial status; benefit rates and average benefit, by risk; average family benefit, by risk, and by family size and composition; average family income of beneficiary families from all sources, by risk, and by family size and composition; social security fund income, by source and by risk; expenditures for social security, by object, by risk.

With respect to other types of data needed, the following may be mentioned: total population, by age, sex, familial status, dependency status; number of persons in the labor force, by employment status, age, sex, familial status; number of persons disabled on an average day, by type of disability (*i.e.*, long or short-term), age, sex, familial status, recent labor force status; average wage, for all employments, or in specified types of employment; minimum wage; average family income from all sources; national income.

Discussion

Mr. Woodbury's remarks on Mr. Cohen's paper, read on his behalf by Miss Rodmer, were developed from the point of view of the problems of assembling and compiling the data:

"... The data on social security fall into three main groups: First, there are the data that form a record of the operations of the social security programs—Mr. Cohen's *absolute* figures... These are indispensable, the raw materials of any national or international statistics. They lack, however, certain elements required to make them significant.

"Secondly, there are the rates and ratios that measure risk and throw light upon the meaning of the several social security measures. In computing these for purposes of international comparison, account has to be taken as Mr. Cohen points out, of differences in the scope and provisions of the various systems. These rates are of different type. ...the scope of the figures must be comparable: This can be obtained, for example, by limiting figures to particular branches of economic activity....

"In the field of morbidity much greater difficulties are found.... The Statistical and Medical Commissions of the Inter-American Committee on Social Security, under the auspices of the International Labor Office, examined this matter with a view to arriving at a standardization of the sickness statistics established by social insurance institutions on this continent....

"The relation between age or sex and morbidity is well known. However, elimination of the influence of differences in age and sex distribution by standardization is not sufficient; rates must actually be given for age and sex groups as the relation is by no means constant....

"The incidence of specific diseases differs from one insured population to the other and a common classification by diagnoses is indispensable for purposes of comparison.

"As in the case of accident statistics, the definition of the units included in each set of morbidity statistics frequently differs. The unit of illness or incapacitating illness recorded may be a spell of illness ending, or one beginning or one current during the period under review. Again, it may be an individual person suffering one or more spells of illness during such period. Common definitions have to be agreed upon. Finally the meaning of the term "persons exposed to risk" to which the units of morbidity must be related if comparison is to be made, should be agreed upon, the time of exposure being generally deemed the most appropriate measurement. As mentioned at a previous meeting, the Statistical and Medical Commissions of the Inter-American Committee on

Social Security have proposed standard rules on these points. They have also set out and defined the data that are deemed desirable in medical care statistics.

"In the field of unemployment the usual common measure is to compare the number of persons unemployed as of a given day with the total of employed and unemployed—the labor force—on the same day. An alternative method is to compare the days of unemployment during a given period, with the total number of possible man-days of work during the same period.... In practice, various modifications are found... There are also differences in the scope of the unemployment insurance systems, which seriously affect comparability....

"In all these calculations, the distinction must be kept in mind between statistics that bear upon the whole problem, such as accidents or unemployment, and those which are specifically limited to social security measures, such as compensated accidents, sickness and unemployment. The latter are often used as a substitute for the former. While the former type is always the more significant for purposes of analysis of the social phenomena with which they deal, the latter are useful with respect to the specific problems of coverage and risk involved in the social security program itself....

Thirdly, the ratios with which Mr. Cohen's paper are specifically concerned deal with appraisal and evaluation. These are of primary importance in policy making and for the appraisal of the social security systems as a whole and in their respective parts. The ratios proposed, however, are difficult to construct, since usually at least one member of each ratio must be estimated.

Each of the different ratios must be considerd by itself. To show overall coverage of a system as a whole, for example, of old age protection, the number of persons covered by old age insurance can be compared to the total population over a given age, or to the wage-earning part of the population, if not an overall coverage but merely the wage earner coverage is the point of concern. Such ratios then can be used to answer the question whether any important sector of the population is not covered, and thus furnish a basis for decision whether modifications of the legislation are needed...

"...The problem, of course, is to obtain such a figure that can be accepted as a satisfactory estimate for purposes of comparison."

Mr. Cohen agreed with his critics that his suggestions are capable only of limited immediate application, but indicated that he hoped for development along the lines he proposed.

Mr. **Stowman** discussed the origin in the League of Nations of work on health indices, and stated that he believes the preliminary and experimental work in this field is bearing fruit. He expressed the hope that the World Health Organization would be able soon to join its efforts with those of the Pan American Sanitary Bureau.

Mr. **Bernstein** discussed his experience with a statistical study of the B.C.G. treatment for tuberculosis, during which a basis in the operation of the experiment rendered invalid the drawing of a positive conclusion from the statistics. Mr. Bernstein argued the great value of laboratory experiments made parallel with public health experiments.

PROBLEMAS DEL DESARROLLO DE LOS INDICES DE SALUD EN AMERICA LATINA

by Manuel de Viado M.D.
Oficina Internacional del Trabajo

La apreciación exacta de las condiciones de salud de una comunidad, estado, región, o país, ha preocupado constantemente a los servicios de medicina pública que han recurrido, como medio de comparación, al uso de las estadísticas vitales. Uso que, como un medio de conocer el estado de salud colectivo, tiene ya una tradición más que centenaria. Sin embargo, a pesar de la antigüedad de su uso y de los muchos trabajos realizados, la comparación de las estadísticas continúa siendo difícil, aumentando las dificultades a medida que, de la comparación regional o nacional, pasamos a una comparación internacional.

La ausencia de datos importantes en algunas comunidades, regiones o países; la diversidad de métodes de compilación o clasificación empleados; la variabilidad de los conceptos básicos de interpretación, la excesiva multiplicidad de organismos interesados en estadísticas virales; la diferencia de intereses y objetivos que animan a cada uno de estos organismos y, a veces, la falta de standards apropiados o la no existencia de una adecuada y moderna legislación estadística, son algunas de las principales dificultades que se encuentran.

Por otra parte, a medida que los grandes avences de la técnica precisan los principios comunes que los permitirán llegar al goce del grado máximo de salud, entendida como "el estado de completo bienestar físico, mental y social," se ha ido comprendiendo la estrecha interdependencia que existe, en materias de salud, entre las diversas comunidades y regiones de un país, así como entre las diversas naciones de un continente y entre los diversos continentes de la tierra. Los problemas sanitarios no son hoy problems locales o regionales sino problemas mundiales. La necesidad de una comparación continental e internacional de las estadistícas vitales sobre bases uniformes se ha hecho así cada día más y más imperiosa.

Es por eso que, sintiendo esa necesidad desde hace ya años, y a pesar de las dificultades, se han realizado muchos y valiosos esfuerzos para obtener una base uniforme de comparación. La contribución de las Conferencias Internacionales de Revision de la Lista de Causas de Muerte; de las Conferencias Sanitarias o de Higiene Pública, panamericanas, europeas y asiáticas; de la Organización de Higiene de la Liga de las Naciones; de la Oficina Internacional del Trabajo; de la Oficina Sanitaria Panamericana; del Instituto Internacional de Estadística; del Instituto Interamericano de Estadística; del Comité Interamericano de Seguridad

Social; de la Asociación Americana de Salud Publica; de los organismos estadísticos nacionales; de las asociaciones profesionales interesadas y de todos los hombres y mujeres que trabajan en los servicios de medicina pública del mundo, no debe ser olvidada, porque todos esos esfuerzos y contribuciones representan pasos hacia la meta común: una buena comparación internacional sobre bases uniformes.

Hace ya tiempo que las naciones estadísticamente más adelantadas nos enseñaron que las estadísticas vitales son la medida del estado de salud de la comunidad, la representación de los resultados obtenidos por la profesión médica y profesiones conexas con la medicina en la defensa de la salud del pueblo y los guías para mejorar el trabajo del sanitarista, iniciar o transformar los planes de accion y solicitar la colaboración pública. Pero las cifras estadísticas representan también algo mas. Son la expresión estática de los diferentes elementos biológicos, económicos, culturales y sociales de cada comunidad, estado, región o país, que influencianúel estado de salud colectivo.

No debe extrañarnos así que, en el curso de los años, a las simples tasas e índices de natalidad, morbilidad y mortalidad, se hayan ido agregando, en el desarrollo histórico, datos sobre las condiciones sanitarias de las escuelas, el estado de la vivienda, el consumo de leche y otros alimentos protectores, el cuidado de la madre y el niño, la atención hospitalaria, el número de profesionales que trabajan en la comunidad y aun el costo de la atención médica por cada individuo atendido, como medios de obtener la más amplia y exacta información posible sobre el estado de de salud la comunidad, región o país.

A medida que la medicina y la estadística adquieren cada día un mayor contenido social, el pensamiento y la acción han pasado gradualmente de la consideración única de los problemas médicos a la consideración de los problemas económicos y sociales. Los índices de salud van así, poco a poco, en relación con el desarrollo de este pensamiento social, transformándose de una simple medida del estado de salud de una comunidad en un medio útil de obtener un retrato de las relaciones entre el estado de salud de los hombres y mujeres de una comunidad, región o, país, y el medio ambiente geográfico, económico y social en que esos hombres y mujeres nacen, viven, trabajan y mueren.

La utilidad de los índices de salud, así comprendidos, está señalada clarament por los efectivos servicios que prestan, en los Estados Unidos de Norte América, los *Health Practice Indices*, desarrollados a lo largo de muchos años de valiosa experiencia por el *Committee on Administrative Practice* de la *American Public Health Association*, con la colaboración económica de importants entidades interesadas en el ciudado de la salud pública, tales como el *Commonwealth Fund* y el *Milbank Memorial Fund.*

Los resultados obtenidos en un plano nacional en los Estados Unidos de Norte América, con los materiales y experiencias de las encuest as sanitarias y los formularios de apreciación, movieron a la Organización de Higiene de la Liga de las Naciones a iniciar el estudio, en 1935, de la possibilidad de una extensión de las estadísticas vitales de higiene pública y de los índices de apreciación empleados, en un plano internacional.

En 1936, Stouman y Falk publicaron en el Boletín Trimestral de la Organización de Higiene de la Liga de las Naciones las primeras proposiciones, objetivas y numéricas, de los índices de vitalidad y salud, condiciones del medio ambiente y actividades administrativas, que consideraban necesarios para un estudio comparado de los diversos factores que influyen en la salud pública de una colectividad.

Este primer trabajo comprendía tres clases de cuestionarios. Uno detallado, con preguntas sobre la población, natalidad, mortalidad infantil, mortalidad maternal, mortinatalidad, mortalidad general, morbilidad; incapacidad, alcoholismo, accidentes y homicidios; condiciones climáticas y topográficas, distribución social y ocupacional del a población, nivel de cultura, condiciones de habitación y nutrición; gastos en salud pública persona que trabaja en los diferent servicios de medicina pública, denuncia, investigación y hospitalización de las enfermedades agudas cumunicables, lucha antivenérea y antituberculosa, higiene maternal, infantil y escolar, condiciones generales sanitarias, control de alimentos, higiene industrial, educación sanitaria, seguros sociales y demás medidas de protección de la salud, etc. Otro abreviado, reducido a 100 índices sobre los tres aspectos principales, salubridad, medio ambiente y actividades de la medicina pública, y, un tercero, más corto, reducido a 60 puntos principales.

Los trabajos, hoy ya clásicos, realizados en New Haven por Stouman con la colaboración de Departamento del Salud Pública de la Universidad de Yale y los organismos estadísticos del Estado de Connecticut y la ciudad de New Haven, probaron la posibilidad de aplicar los cuestionarios de los Indices de Salud a una comunidad norteamericana. Más tarde, la publicación, por el mismo Stouman, de índices de salud para el distrito rural de Mezökövesd en Hungría en 1937, para la ciudad industrial de Bruselas en 1938 y los estudios realizados por Kacprzak y Tomcsik en otros distritos rurales de Polonia y Hungría, probaron las posibilidades de su uso en las regiones rurales, tanto como en las regiones industriales, y la necesidad de simplificarlos, si debían llegar a tener un carácter internacional. En 1938 un grupo de expertos seleccionó una serie de índices aplicables a la zonas rurales y en octubre del mismo año un Comité de Indices de Salud de la Organización de Higiene de la Liga de las Naciones, compuesto por representantes de los Estados

Unidos, Polonia, Gran Bretaña, Hungría, Suecia y Francia, fijó 52 proposiciones para un standard internacional que contemplaba los tres aspectos principales (índices de vitalidad y salud, índices del medio ambiente e índices de acción pública) establecidos anteriormente por Stouman y Falk, mientras en 1939 Selleslags aplicaba los índices rurales al estudio de la provincia belga de Luxemburgo.

La finalidad de los índices de salud así establecidos era el estudio global de la salud pública en sus tres grandes aspectos; la salud y vitalidad de los individuos, los factores del medio ambiente que los rodean y los procedimientos de lucha sanitaria utilizados. Las preguntas sobre demografía, mortalidad, morbilidad, metereología, habitación, saneamiento, consumo, comunicaciones, instrucción, personal y armamento sanitario, actividades médico sociales y de asistencia social, por el carácter mixto que ahora tenían eran aplicables tanto a los medios urbanos como a los medios rurales.

La Segunda Guerra Mundial hizo que en la Europa ocupada, no fuera propicio el ambiente para continuar desarrollando los índices de salud, aunque hay referencias a ellos en algunos trabajos de carácter general, como el publicado por Y. Biraud, hoy Secretario General Adjunto de la Organización Mundial de la Salud, sobre el estado sanitario de Europa en 1944. Es en el continente americano donde encontramos la continuidad de desarrollo de la aplicación de los índices de salud. Los Estados Unidos, a pesar de su gran participación en el esfuerzo bélico mundial, continuaron mejorando sus *Health Practice Indices* y América Latina ha realizado diversos estudios desde 1940.

En ese año la Delegación del Brasil a la IV Conferencia Panamericana de Directores Nacionales de Sanidad, reunida en Washington, presentó un formulario de nuevos índices sanitarios a la consideración la comisión especial de dicha conferencia. Estos nuevos índices sanitarios panamericanos, conservan el esquema general de Stouman y Falk, amplían algunos rubros y suprimen otros, para hacerlos más adaptables a los países latinoamericanos. Constan de 72 preguntas, algunas no muy explícitas, refiriéndose 18 de ellas a los índices de vitalidad y salud, 17, a los índices relativos al medio ambiente y 37 a las actividades de los servicios sanitarios.

La moción del Brasil fué acogida favorablemente en los términos siguientes; "La IV Conferencia Panamericana de Directores Nacionales de Sanidad, aprueba los índices sanitarios mínimos que constan en las actas, y recomienda a las autoridades sanitarias, que sean aplicadas por vía de enseyo en una o dos poblaciones de cada una de las Republicas Americanas, y que los resultados obtenidos sean presentados a la Oficina Sanitaria Panamericana."

De acuerdo con esa recomendación se han publicado ya datos para diversas ciudades de Brasil, Cuba, Chile, México, Nicaragua, Paraguay, Peru y Venezuela.

En el cuadro que se acompaña resumimos las respuestas a las preguntas del cuestionario de los índices de salud panamericanos.

Qué nos enseña, en grandes líneas, la experiencia latinoamericana de la aplicación de los índices de salud? Como resultado favorable, en primer lugar, su efectiva utilidad como complemento de las estadísticas vitales, en cuanto nos proporcionan mayores elementos de información acerca del medio ambiente y de la acción de los servicios de medicina pública, si ellos son bien aplicados.

En ese sentido, los primeros índices de salud, según los esquemas de Stouman y Falk modificados, son de mayor utilidad—por considerar un mayor número de factores economicos y sociales—que los índices sanitarios panamericanos, des carácter más restringido por ser más sanitarista. Así se explica que en las respuestas a dichos índices no se proporcionen datos que existen en esos mismos países, pero que están en posesión de otros servicios tales como los del seguro social, educación, direcciones generales de estadística, hospitales, etc.

Este punto es de importancia, como experiencia, porque ningún índice de salud reflejará las condiciones verdaderas del estado de salud de una localidad, region país—para fines de comparación—si no considera debidamente todos los factores que actuan sobre la salud de la población, así como todos los servicios, sin distinción de funciones, a través de los cuales se protege la salud del pueblo.

Esta necesidad de una consideración general del problema se hace cada vez más necesaria, a medida que una mejor comprensión de lo que debe ser y será la medicina de nuestros tiempos hace desaparecer las divisiones entre medicina curativa y medicina preventiva, medicina individual y medicina colectiva, para no hablar sino de una unidad orgánica; la medicina social.

Los indices de salud, adaptados a las particulares necesidades de cada estudio, son aptos para estudios de medicina social cuando consideran debidamente todos los factores en ella implicados. Los trabajos realizados en Brasil, en un análisis general do los servicios médicos del país, y los realizados en Chile por el Departamento de Estadística de la Dirección General de Beneficencia y Asistencia Social, en relación con la situación de los hospitales del país o por la Caja de Seguro Obligatorio, en relación con la labor realizada en el seguro social, usando los índices europeos modificados, son una prueba de la adaptabilidad de dichos índices y de la conveniencia de su empleo como medio de encuesta previa a toda acción médica planificada.

PAIS	México	Peru	Peru	Peru	Peru	Peru	Venezuela	Venezuela	Venezuela	Venezuela
Ciudad Fecha del Indice	México 1940	Lima 1940	Callao 1943	Arequipa 1943	Truji 1943	Iquitos 1943	Caracas 1943	Maracaibo 1943	La Guayra 1942	Pto. Cabello 1942
1. Poblacion (calculo)	1,461,410	520,528	87.936	82.085	48.309	38,000	354,970	143,644	35,245	22,254
2. Coeficiente anual de crecimiento (por 1.000)	8.10	13.06	18.6	13.6	12.9	40.9	66.9	36.7	39.0	14.3
3. Distribucion etarea;										
a) menores de 15 años	32.80	29.78	37.7	39.2	42.2	44.0	30.8	36.1	32.9	29.1
b) de 45 a 60 años	9.57	9.48	13.1	13.2	45.8	10.7	15.4	14.6	15.8	16.5
c) mayores de 60 años	4.72	0.02	—	—	—	—	—	—	—	—
4. Tasa de natalidad (por 1,000 mujeres de 15 a 44 años)	129.8	114.0	164.8	123.2	149.6	221.0	148.9	137.0	175.0	124.9
5. Mortalidad neonatal (en el 1er. mes por 1,000 nacidos vivos)	53.0	22.70	20.70	28.3	14.5	30.5	25.8	20.4	12.4	36.8
6. Mortalidad infantil (por 1.000 nacidos vivos)	97.6	113.19	80.5	96.3	60.4	89.0	69.0	103.8	64.1	92.2
7. Mortalidad por diarrea (en el 1er. año por 1,000 nacidos vivos)	43.8	38.23	24.6	31.8	14.0	16.6	22.7	47.3	31.2	25.4
8. Mortalidad maternal (por 1,000 nacidos vivos)	4.0	3.91	1.7	6.3	1.7	2.6	2.9	4.6	4.0	2.8
9. Tasa de mortalidad general (por 1,000 habitantes)	22.5	16.34	17.0	18.6	17.5	23.2	13.5	15.6	12.7	21.2
10. Tasas especificas de mortalidad (por 100,000 habitantes) en;										
a) Tifoidea y paratofoidea	19.5	12.85	17.1	29.2	22.8	2.6	11.0	7.0	14.2	9.0
b) Disenterias	7.6	—	3.4	8.5	2.1	16.0	11.6	16.0	8.5	13.5
c) Tuberculosis	81.2	268.98	506.0	490.0	426.4	223.0	237.2	204.7	238.3	464.7
d) Difteria (por 100,000 manores de 15 años)	18.7	25.6	14.0	12.4	59.0	5.9	11.0	17.4	0.0	30.9
e) Malaria	2.8	24.94	29.6	—	24.8	50.0	5.6	64.0	11.3	63.2
f) Sarampion (por 100,000 menores de 15 años)	81.4	34.17	0.0	3.1	9.8	89.8	5.5	9.6	0.0	0.0
g) Cancer (por 100,000 mayores de 45 años)	121.8	545.50	530.3	303.4	301.8	609.1	563.4	448.2	395.0	463.0
h) Enfermedades del corazon (igual que cancer)	57.8	99.76	681.8	533.3	729.4	341.1	865.3	662.3	502.8	762.5
i) Infecciones respiratorias	590.7	326.91	*—	—	—	—	—	—	—	—
j) Viruela	0.21	0.9	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0
k) Suicidios	0.8	6.52	—	—	—	—	—	—	—	—
l) Accidentes automovilisticos	14.00	12.66	—	—	—	—	—	—	—	—
11. Porcentaje de seroreacciones positivas (sifilis)										
a) Mujeres gravidas	—	12.49	14.8	15.8	—	—	8.0	9.0	15.0	22.2
b) Grupos de poblacion	—	—	—	—	12.4	—	14.0	9.5	14.0	30.0
12. Porcentaje de tuberculosis en los grupos de examen toracico serados	—	del 60.2 al 90.4	10.1	7.4	3.9	—	del 2.2 al 12.9	del 2.2 al 8.7	—	del 1.4 al 6.1
13. Numero de leprosos										
a) en leprosarios	188	10	no hay	no hay	no hay	350.0	20.0	700.0	16.0	no hay
b) a domicilio	73	—				65.0	14.0	0.0	0.0	
14. Indice endémico de Ross	—	—	—	—	—	—	—	—	—	—

PAIS	Venezuela	Venezuela	Venezuala	Venezuela	Chile	Chile	Chile	Nicaragua	Nicaragua	Nicaragua	Paraguay	Paraguay	Paraguay	Paraguay	Cuba.
Ciudad Fecha del Indice	Valencia 1942	Barquisamet 1942	S. Cristobal 1942	Maracay 1943	Santiago 1943	Concepcion 1943	Antofagas 1943	Managua 1943	Leon 1943	Granada 1943	Asuncion 1943	Villarrica 1943	Concepcion 1943	Encarnacion 1943	Habana 1943
1.	54,586	56,206	32,417	31.405	993.310	85.248	46.438	88,519	50,069	35,202	121.202	16.124	16,124	16,155	676,376
2.	19.4	98.5	88.5	22.0	18.4	11.5	12.0	—	—	—	3.45	1.70	3.49	4.0	11.3
3.															
a)	33.8	34.0	36.5	38.0	30.33	31.71	32.81	—	—	—	—	—	—	—	19.9
b)	17.1	13.3	13.7	13.3	16.93	15.57	19.11	—	—	—	—	—	—	—	16.2
c)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
4.	108.0	109.8	160.7	129.7	124.74	170.2	140.2	57.07	38.15	41.46	22.2	38.28	45.34	57.0	51.8
								Por 1,000	habitantes			Por 1,000	habitantes		
5	37.3	48.4	42.3	34.6	34.57	65.69	38.65	25.93	12.20	33.84	23.23	42.62	16.0	29.0	—
6.	109.0	140.6	122.6	99.7	123.3	207.12	113.5	140.54	75.29	182.32	61.62	61.62	34.15	46.0	—
7.	24.7	28.9	40.9	24.7	20.6	52.79	4.98	47.90	51.0	60.77	10.0	—	—	—	—
8.	10.2	6.1	16.5	6.9	7.4	5.0	3.7	4.35	6.89	4.83	3.34	7.94	3.82	2.0	—
9.	22.7	18.9	23.4	17.7	18.4	29.39	18.11	26.06	13.15	21.65	12.16	22.46	11.44	19.0	14.3
10															
a)	14.6	5.3	37.0	12.7	2.15	5.87	2.15	40.65	12.13	80.17	15.58	12.0	24.0	12.0	23.4
b)	11.0	10.7	30.8	6.4	2.5	1.17	0.0	65.85	12.13	45.82	8.87	6.0	19.0	141.0	3.6
c)	363.4	227.7	283.8	210.2	336.62	519.66	340.23	89.43	26.30	45.82	10.29	36.0	129.0	80.0	189.2
d)	5.6	44.8	16.9	46.8	13.24	7.39	19.69	0.81	2.02	—	—	—	6.6	6.0	17.8
e	294.9	42.7	138.8	35.0	no hay	no hay	no hay	322.75	175.95	395.16	2.3	12.0	6.0	37.0	1.5
f)	0.0	15.0	101.4	9.4	0.5	55.48	0.0	—	—	—	—	—	—	—	22.3
g)	407.1	374.6	630.5	454.9	553.59	648.13	360.52	5.69	10.11	42.95	67.77	6.0	68.0	37.0	940.9
h)	846.4	548.5	427.8	790.0	1,495.29	1,198.28	1 475.89	85.36	68.76	85.91	15.56	66.0	284.0	241.0	1.489.8
i)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
j)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	—	—	—	—	—	0.0
k)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
11.															
a)	21.6	27.0	15.8	16.6	8.28	2.96	9.2	—	—	—	—	—	—	—	del 2.4 al 3.4
b)	18.2	16.0	23.0	24.0	21.73	17.81	18.7	—	—	—	—	—	—	—	13.0
12.	3.7	7.0	del 2.3 al 2.9	1.9	2.0	—	—	—	—	—	3.36	—	—	—	—
13.															
a)	1.0	0.0	3.0	0	no hay	no hay	no hay	—	—	—	400	—	—	—	—
b)	0.0	8.0	—	0				—	—	—	—	—	4	5	—
14.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Ciudad	México	Lima	Callao	Arequipa	Trujillo	Iquitos	Caracas	Maracaibo	La Guayra	Pto. Cabello
15. Porcentaje de escolares con Anquilostoma	no hay	no hay	no hay	no hay		6	4.0	0.7	0.0	—
16. Porcentaje de escolares sin defectos; a) dientes;	94.2	—	—	—	—	80	98	97	—	—
b) vista	91.1	—	—	—	—	—	7.6	11.8	—	—
17. Coeficiente de alienados internados (por 100,000)	18.0	—	—	—	—	—	211.3	346.0	0.0	0.0
18. Coeficiente de victimas de accidentes de automovil (por 100,000 habitantes)	164	201.25	52.3	81.6	58.0	81.6	36.6	38.3	36.9	40.4
19. Temperatura media										
a) Maxima	27.3	24.2	27.0	25.0	32.9	37	28	34.8	—	32.
b) minima	2 3	16.76	14.6	0.0	9.4	15	13.3	22.6	—	21
20. Luz solar	74.5	34.3	(datos en horas)		(datos en horas)		54.0	61.6	—	—
21. Precipitacion pluvial del mes mas lluvioso	179.4	71.0	12.0	187.7	16.3	181	820	514.3	467.6	48.9
22. Densidad de poblacion urbana por cuadrado	10,930	8,157	—	—	—	9,500	252.3	—	—	—
23. Densidad por habitacion	0.76	—	5.0	5.0	5.0	5.0	2.1	2.1	2.2	2.4
24. Porcentaje de la poblacion masculina que trabaja en industrias	18.0	12.9	—	—	—	—	—	—	—	—
25. Porcentaje de mayores de 15 años que trabajan: hombres	74.7	77.6	—	—	—	—	—	—	—	—
mujeres	10.3	29.5	—	—	—	—	—	—	—	—
26. Porcentaje de menores de 15 años que trabajan	24.0	3.51	—	—	—	—	—	—	—	—
27. Numero de empleados domésticos por 1,000 habitantes	49.8	6.47	—	—	—	—	—	—	—	—
28. Numero de automoviles por 1,000 habitantes	3.1	2.0	—	—	—	—	—	—	—	—
29. Porcentaje de analfabetos menores de 15 años	11.34	15.48	4.3	12	11.4	17.5	18.2	24.8	27.8	30.2
30. Porcentaje de asistencia escolar (menores de 15)	62	—	58.1	60.2	67.4	62.4	56.0	70.1	71	69.1
31. Porcentaje de la poblacion que vive en habitaciones colectivas	0.6	—	—	—	—	—	—	—	—	—
32. Porcentaje de casas con una sola familia	14.2	—	—	—	—	—	—	—	—	—
33. Promedio de aposentos por vivienda	3.6	—	—	—	—	—	—	—	—	—
34. Porcentaje de viviendas sin baño privado	55.6	—	—	—	—	—	—	—	—	—
35. Consumo de leche por cabeza	54	40	litros anuales		—	por dia	224 cc	12 cc	47 cc	86 cc
36. Numero de médicos (por 100,000 habitantes)	117.5	121.25	45.5	70.7 —	47.6	65.8	132	49.3	51.4	62.9
37. Numero de enfermeras y vivitadoras sociales	18.0	—	—	—	—	—	31	4.5	17.1	22.5
38. Porcentaje de certificados de defuncion firmados por médico	100	100	95	95	. 90	97.9	100	100	100	100

Ciudad		Valencia	Barguisamet	S.Cristobal	Maracay	Santiago	Concepcion	Antofagas.	Managua	Leon	Granada	Asuncion	Villarica	Concepcion	Encarnacion	Haban
15.		del 14 al 40	—	11.5	—	0.0	0.0	0.0	—	—	—	4.44	del 35 al 45	del 40 al 75	del 50 al 70	2.9
16.	a)	90	—	—	74.0	70.0	—	—	—	—	—	—	—	—	—	96
	b)	5	—	—	—	16.0	—	—	—	—	—	—	—	—	—	21
17		183.2	195.0	0.0	0.0	—	—	—	—	—	—	140	—	—	—	—
18.		45.8	51.6	55.5	35.0	224.3	60.99	58.14	—	—	—	51	—	—	—	60.3
19.	a)	31.6	30.7	23.3	32.4	30.2	25.0	23.4	31.5	—	—	38	37.6	39	—	31.6
	b)	19.9	19.2	20.5	13.8	3.3	6.0	10.7	27.7	—	—	4	6	10	—	17.7
20.		—	—	—	60.0	datos en dias en Chile			—	—	—	—	—	—	—	—
21.		164.5	558.1	1,375.0	977.3	361.0	1,305.0	9.04	160.0	—	—	—	—	—	—	68.9
22.		—	5,000	—	—	959.7	—	—	—	—	—	—	—	—	—	5,985.6
23.		1.95	2.4	2.03	2.1	—	—	—	—	—	—	—	—	—	—	—
24.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
25.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
26.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
27.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
28.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
29.		25.5	39.7	62.8	30.9	9.98	—	—	—	—	—	—	—	—	—	21.4
30.		66.8	55.2	27.5	64.3	75.	—	—	—	—	—	80	60	70	80	85
31.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
32.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
33.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
34.		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
35		83 cc	100 cc	62.5 cc	10 cc por dia		—	—	200 cc	—	—	—	—	—	—	89.06 cc
36.		73.3	61	55.0	76.4	11.3	96.19	58.14	67.55	95.88	76.59	44.9	24	30	18	297.2
37.		18.3	19	31.0	35	—	18.77	23.69	12	—	—	32	25	43	30	18.5
38.		100	.100	100	100	100.00	100	100	85.18	65.38	87.17	94.18	35	98.91	67	100

Ciudad	México	Lima	Callao	Arequipa	Trujillo	Iquitos	Caracas	Maracalbo	La Guayra	Pto.Cabello
39. Porcentaje de muertes por causas desconocidas	0.7	3.13	5.2	1.2	13.1	14.7	1.5	11.7	5.1	4.7
40. Examenes de laboratorio por caso de tofoidea	1.0	—	—	—	—	—	—	—	—	—
41. Examenes de leche por 1,000 litros consumidos	1.5	7.0	—	—	—	—	—	—	—	—
42. Examenes de agua potable (por semana)	14.0	21.0	—	—	—	—	—	—	—	—
43. Porcentaje de muestras (de acuerdo con standards)	18.0	—	—	—	—	—	—	—	—	—
44. Numero de casos conocidos por muerte de; a) tifoidea	2.7	8.8	15.0	27	11	1	6.7	5.0	7	6.5
b) disenteria	0.4	—	3	7	1	6	3.0	45.8	2.7	41.2
c) tuberculosis	17.2	—	445	403	206	85	1.7 nuevos	1.3 nuevos	—	1.1nuevos
d) difteria	4.7	7.2	—	—	—	—	—	—	—	—
e) malaria	0.9	17.2	—	—	—	—	—	—	—	—
f) sarampion	4.5	—	—	—	—	—	—	—	—	—
45. Inmunizaciones contra; a) viruela	312,660	79,403	10,576	33,342	9,683	1,233	49,204	12,189	2,737	1,703
b) tifoidea	997	4.219	1,647	21,382	2.664	0	3.819	6,514	4,318	710
c) difteria	71,950	—	30	509	10	0	7,916	3,318	113	136
d) tuberculosis	—	851	—	—	—	—	—	—	—	—
46. Numero de casos en	3.786	351	264	160	83	100	3.209	669	387	676
tratamiento de; sifilis	—	—	2,491	5,843	1,179	425	—	1,577	—	—
blenorragia	9.130	2.815	701	138	123	346	2.860	806	—	289
47. Casos de contacto por caso de lepra	3	0	no hay	no hay	no hay					no hay
48. Consultas por cada tuberculoso registrado	1.07	—	—	—	—	—	—	—	—	—
49. Casos de contactos por cada caso de tuberculosis	1.13	0.9	—	—	—	—	—	—	—	—
50. Visitas de enfermeras por cada caso de tuberculosis	3.08	1.2	—	—	—	—	—	—	—	—
51. Porcentaje de enfermos no registrados sobre el total de defunciones por tuberculosis	—	—	—	—	—	—	—	—	—	—
52. Mortalidad por tuberculosis en 1,000 casos registrados por mas de 6 meses	—	—	—	—	—	—	—	—	—	—
53. Porcentaje de mujeres gravidas por cada 100 nacimientos inscritos	27	74	—	—	—	—	—	—	—	—
54. Porcentaje de partos ocurridos en maternidades	44	52.6	—	—	Por médicos	—	71.9	40	27.7	32
55. Número de niños registrados por 100 nacidos vivos	58.5	99.5	—	—	—	—	—	—	—	—
56. Numero de niños que rettben leche gratuitamente por 100 nacidos vivos	1.3	51.6	—	—	—	—	—	—	—	—
57. Porcentaje de niñon menores de 5 años registrados	17.6	15.06	—	—	—	—	—	—	—	—
58. Porcentaje de escolares examinados	100	—	—	—	—	—	—	—	—	—

Ciudad	Valencia	Barquisamet	S. Cristobal	Maracay	Santiago	Concepcion	Antofagas.	Managua	Leon	Granada	Asuncion	Villarica	Concepcion	Encarnacion	Habana
39.	18.2	29.5	9.6	13.1	3.32	0.4	5.18	0.13	0.77	0.40	4.88	65	1.09	0.94	0.4
40.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
41.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
42.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
43.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
44. a)	4.9	5.7	2.9	6.2	5.8	3.2	29	—	—	—	23	2	4	2	—
b)	1.0	7.8	13.5	1.5	1.3	1	0	—	—	—	13	1	3	23	—
c)	—	3.5 nuevos	2.7 nuevos	2.3 nuevos	—	—	—	—	—	—	152	6	21	13	—
d)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
e)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
f	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
45. a)	3,418	4,001	2.674	2,521	121,050	20.715	11.257	—	—	—	31	19	40	52	40.9
b)	3,045	513	3.149	1,172	63,177	2,364	00	—	—	—	41	19	45	60	78.9
c)	474	822	1,368	323	2,199	447	00	—	—	—	—	—	—	—	57.9
d\	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
46	1.652	77	632	70	11.957	1.323	929	Ley de medicina preventiva chilena			1.742	297	320	—	—
	3,808	—	—	494	—	—	—	—	—	—	—	—	—	—	—
	280	127	442	201	6,088	1,688	366				1.316	13	115	—	—
47.	—	—	—	0	no hay	no hay	no hay	—	—	—	—	—	—	—	—
48.	—	—	—	—	hay datos en estadisticas del Seguro Social			—	—	—	—	—	—	—	—
49.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
50.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
51.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
52.	—	—	—	—	Existen datos en el Seguro Social Direccion General de Beneficencia, Servicio Nacional de Salubridad y otras instituciones.			—	—	—	—	—	—	—	—
53	—	—	—	—				—	—	—	—	—	—	—	—
54.	32.9	38	—	—	—	—	—	—	—	—	13	8.5	27.4	19.6 por médicos	
55.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
56.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
57.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
58.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Cuidad	México	Lima	Callao	Asequipa	Trujillo	Iquitos	Caracas	Maracaibo	La Guayra	Pto.Cabello
59. Porcentaje de escolares que reciben alimentacion escolar	—	—		—	—	—	—	—	—	—
60. Numero de horas semanales de ejercicios fisicos en el programa escolar	3	1	—	—	—	—	—	—	—	—
61. Porcentaje de cases conectadas con la red de abastecimiento de agua potable	47.6	83	95	90	90	no hay	60	60.2	78	55
62. Porcentaje de casas no conectadas con la red, pero con pozos sin contaninacion	38	16.9	—	—	—	100	—	—	—	—
63. Porcentaje de casas conectadas con la red de desague	69.8	78.2	95	90	90	80	—	15	58	13
64. Porcentaje de casas no conectadas, con pozos sépticos	14.5	—	5	10	10	no hay	—	47	14.1	30
65. Frecuencia del acarreo de basuras	diario	diario	diario	diario	diario	diario	cada 2 dias	diario	2v.p.semana	semanal
66. Porcentaje de leche pasteurizada dada la poblacion	46	11	—	—	no hay	no hay	50	9.3	no hay	no hay
67. Inspecciones del abasto de leche	312 anuales	—	—	—	—	—	—	—	—	—
68. Inspecciones de los reaturantes	si	si	no	no	no	no	si	si	si	si
69. Inspeccoines de las fabricas	si	si	no	no	no	no	no	no	no	no
70. Numero de camas de hospital por 1.000 habitantes	4.88	5.4	9.7	7.1	6.4	6.6	7.5	8.7	1.9	7.5
generales	1.78	4	—	—	—	—	2.6	4.6	1.5	6.3
infecciosas	0.2	0.3	—	—	—	—	0.3	—	—	—
mantales	2.25	—	—	—	—	—	2.0	3.5	—	—
tuberculosis	0.44	0.6	—	—	—	—	2.0	—	—	—
maternidad	0.21	0.5	—	—	—	—	0.6	0.6	0.4	1.2
71. Gasto per capita en los servicios	0.93	—	14	4	3.4	4	—	—	—	6
curativos y de salud publica	1.59	—	—	—	—	—	6	2.6	5.3	12.04
72. Porcentaje de preguntas contestadas.	91	67	46	46	51	51	56	56	54	54
73. Fecha de publicacion del indice	Oct. 1942	Ago. 1943	Ene. 1946	Ene. 1946	Enero. 1946	Enero. 1946	Abril 1945	Abril 1945	Abril 1945	Abril 1945

Los indices sanitarios han sido publicados en el Boletin de la Oficina Sanitaria Panamericana y la responsabilidad de ellos corresponde a los señores Mario Leon Ugarte y O. Téllez del Peru: Dario Curiel de Venezuela: Jerjes Vildosola de Chile: Jesus Castillo de Nicaragua: y O. Martinez y R. Calvo de Cuba, no citandose los nombres de los autores de los de México y Paraguay.

Ciudad	Valencia	Barquisamet	S. Cristobal	Maracay	Santiago	Concepcion	Autofagas.	Managua	Leon	Granada	Asuncion	Villarrica	Concepcion	Encarnacion	Habana
59.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
60.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
61.	49.5	48	85	43.4	100	95	52.2	60.5	40	40	datos generales			—	92.7
62.	—	—	0	32.1	no hay	5	—	—	—	—	21	—	—	—	57
63.	15.5	20	60	48.1	95	—	—	40	no hay	no hay	21	—	—	—	57
64.	6.2	18	0	32.1	5	—	—	53	75	—	36	—	—	—	3
65.	cada 2 dias	semanal	diario	diario	diario	diario	c.2 dias	c.2 dias	c.2 dias	c.dos dias	si	si	si	si	diario
66.	100	no hay	no hay	no hay	—	—	—	no hay	no hay	no hay	8	—	—	—	18.5
67.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
68.	si	si	si	si	si	si	si	si	si	si	si	si	si	si	si
69.	no	no	no	no	si	si	si	no	no	no	si	si	si	si	si
70.	—	6.1	6	4.2	7.31	9.6	12.9	—	—	—	5.76	4.8	8	8	—
	5.5	4	4.3	3.0	2.2	—	—	—	—	—	—	—	—	—	—
	0.1	—	—	0.6	0.11	—	—	—	—	—	—	—	—	—	—
	1.8	—	—	—	2.82	—	—	—	—	—	—	—	—	—	—
	1.5	1.1	0.6	0.6	0.64	—	—	—	—	—	—	—	—	—	—
	0.5	1	1.1	0.6	0.82	—	—	—	—	—	—	—	—	—	—
71.	—	12	9.81	7.26	—	—	—	—	—	—	—	—	—	—	—
	3.4	3.6	5.49	4.11	195.4	195.4	195.4	—	—	—	—	—	—	—	—
72.	55	54	54	56	55	46	46	34	28	28	43	40	34	34	40
73.	Abril 1945	Abril 1945	Abril 1945	Abril 1945	Mayo 1945	Mayo 1945	Mayo 1945	Sept. 1945	Sept. 194	Sept. 1945	Feb. 1946	Feb. 1946	Feb. 1946	Feb. 1946	Marzo 1947

Los índices de salud son así, a través de la experiencia americana, útiles como medio práctico de comparación y de análisis de la situación de una comunidad, como medio de planificar una acción dica dirigida hacia objectivos determinados y como método de valoración de costos y resultados del trabajo médico y de las profesiones conexas.

Al lado de estas ventajas, la experiencia americana, nos muestra también algunas de sus deficiencias de aplicación. Muchas de las preguntas no han podido ser contestadas por la diferencia de condiciones entre los diversos países de América Latina o por no aplicarse al ambiente regional en estudio. Otras, por no existir datos sobre esas materias en los organismos estadísticos nacionales o en la institución, generalmente la Sanidad Pública, a que pertenecían los autores (aunque, como hemos dicho puedan existir en otros). Varias respuestas son diferentes, porque su falta de precision, deja demasiado juego a la iniciativa personal, o porque cada país puede usar diferentes unidades y métodes de trabajo. La periodicidad y aún el tipo de índice son diferentes, porque han cambiado durante los años transcurridos, por falta de información mas completa de lo realizado en otros países o por seguir demasiado fielmente las publicaciones anteriores efectuadas. Sin contar que parecería necesario considerar en el cuestionario datos acerca de la medicina industrial, la educación sanitaria de la población, los regímemes de seguro social y otros muchos puntos de interés.

Podríamos multiplicar las críticas posibles, pero ellas no son el objeto de este breve esbozo. Basta decir que las deficiencias hacen perder a los resultados mucho de su valor objetivo, aunque persista su valor subjetivo.

Pero sobre todo, los diferentes trabajos realizados y el desarrollo de la aplicación de los índices de salud, permiten desprender los principales problemas del momento actual.

Parece ser el primero de ellos el logro de un índice tal, o de varios índices si es necesario, que sea aplicable universalmente, contemplando las variadas condiciones de vida de los países metropolitanos y de los territorios no metropolitanos, de las naciones o regiones de alta industrialización y de los países o regiones de escaso desarrollo industrial, de las poblaciones urbanas y de las poblaciones rurales y de los diferentes componentes étnico-geográficos, culturales y sanitarios de las diversas comunidades del mundo. En este sentido una revisión del texto standard de los índices de salud del Comité de Indices de preguerra, parecería ser necesaria para poner al día los cuestionarios de los índices de salud, con los adelantos de la medicina y la estadística y con las experiencias obtenidas en su aplicación, en los ultimos años.

Las preguntas que se hagan deberían contemplar, en la forma y cantidad adecuadas, el mayor número posible de elementos económicos y sociales que obran sobre la salud y una serie de factores nuevos de interés para la medicina social, tales como los que se refieren, entre otros, a los problemas del trabajo y a la medicina industrial, a la medicina del seguro social, a la rehabilitación y reeducación profesional, al servicio social y la enfermería sanitaria, a la enseñanza médica y condiciones de trabajo de los profesionales de la medicina, etc., a todas las partes, en fin, especiales pero no independientes, que constituyen la medicina social de hoy y de mañana.

Para la comparación internacional, la periodicidad de elaboración de los índices de salud y si es posible, su simultaneidad en el tiempo, debería ser fijada, sin perjuicio de que cada país o institución interesados pueda efectuarlos, para sus propios fines, cuando lo estime conveniente. En este sentido, los recientes estudios y gestiones efectuados por al Instituto Interamericano de Estadística, en relación con el Censo Continental de las Américas en 1950, podría ser un ejemplo de la posibilidad de lograr, en el plano continental o internacional una comparación uniforme periódica del estado de sauld de las poblaciones, a través de índices de salud comparables.

Para lograr la indispensable uniformidad de métodos y de interpretación, los índices de salud, no deberían ser hechos individualmente, sino bajo una responsabilidad oficial. Esta responsabilidad debería ser, en cada comunidad, región o país, de aquella institución estadística oficial, nacional o local, que—en colaboración con los otros organisoms estadísticos del país, región o comunidad—pudiera darle a las respuestas del cuestionario la garantía e importancía que requiere una contestación completa y plena a todas sus preguntas con un solo criterio y con el empleo del máximo de recursos, de tal manera que su objetivo informador y planificador pueda ser logrado.

Se plantea así indirectamente el problema del punto focal nacional para la ejecución e intercambio de estadísticas vitales.

Si estos principales problemas de los índices de salud, universalidad, definición de objetivos y propósitos, cantidad de índices a usar, preguntas que deben contener, uniformidad de métodos de recopilación e interpretación, unidades a emplear, periodicidad, institución o instituciones responsables de su ejecución e información, las trasladamos a un plano internacional, vemos sus dificultades. Sin embargo, estas dificultades pueden ser vencidas, por los expertos bioestadísticos de los diferentes ramos de la medicina pública con el respaldo de una organización internacional.

Este hecho plantea la pregunta: a quién corresponde ocuparse internacionalmente de los índices de salud? Creemos que, en el momento actual, con los precedentes que ya existen, tal pregunta puede ser contestada porque se ha reconocido plenamente que la solución de ciertos problemas en el campo de la salud depende de una organización internacional. Esa organización existe ya, aunque no todavía legalizada totalmente, y es la Organización Mundial de la Salud de las Naciones Unidas.

Correspondería así, segun nuestra opinion, a la Organización Mundial de la Salud, en colaboración con las otras agencias especializadas de las Naciones Unidas y organismos interesados, el ocuparse de los estudios tendientes al establecimien to de índices de salud uniformes para la comparación internacional. La Organización Mundial de la Salud, puede, naturalmente, si ella lo estima conveniente, obrar a través de los organismos de colaboración regional en sus relaciones nacionales.

Es así como, en este ligero esbozo de los principales problemas del desarrollo de los índices de salud con la experiencia directa de su aplicación en el continente americano, llegamos a la conclusión de su utilidad practica como instrumento de trabajo de la medicina pública, y a la necesidad de encarar la solución de sus principales problemas en un plano internacional a cargo de una organización internacional representativa.

Si tal fuera también el pensamiento de los participantes en estas Conferencias Internacionales de Estadística y las cuestiones aquí esbozadas llegaran a constituir un motivo para más cuidadosos análisis posteriores de los expertos, nuestro pequeño trabajo habría cumplido las finalidades que nos movieron a escribirlo.

BIBLIOGRAFIA

1. American Public Health Association, *Health Practice Indices*, 1943-44.
2. *Arquivos de Higiene do Brasil*, Agosto e Dexembro 1943.
3. Y. Biraud, "L'état sanitaire de l'Europe," *Bulletin de l'organisation d'hygiene*, Société dés Nations, Vol. X, No. 4, 1943-1944.
4. *Boletin de la Oficina Sanitaria Panamericana*, Vol. 21, No. 10, 1942; Vol. 22, No. 8, 1943; Vol. 24; Nos. 4, 5, 9, 1945; Vol. 25, Nos. 1 y 2, 1946; Vol. 26, No. 3, 1947.
5. *Boletín Médico Social*, Caja de Seguro Obligatorio, Años 1945 y 1946.
6. Committee on Health Indices, "Skeleton Standard on the State of Health," *Bulletin of the Health Organization*, League of Nations, Vol. VIII, Nos. 1, 2, 1939.
7. Manuel de Viado, Carlos Solar y Col, *Encuesta médico-social para* 1944. Departamento de Estadistica, Direccion General de Beneficencia y Asistencia Social, Chile, 1945.
8. F. E. Linder, "International Vital Statistics of the Future," *American Journal of Public Health*, July 1944.
9. Mario Leon Ugarte, "Los Indices Sanitarios," *Informaciones Sociales*, Año VI, No. 4, 1942.

10. K. Stouman y J. S. Falk, "Health Indices, A Study of Objectives Indices of Health in Relation to Environnement and Sanitation," *Bulletin of the Health Organización*, League of Nations, Vol. V, No. 4, 1936.

11. K. Stouman, "Health Indices for the City of New Haven," *Bulletin of the Health Organization*, League of Nations, Vol. V, No 4, 1936.

12. K. Stouman, "Health Indices in a Rural Districs of Hungary," *Bulletin of the Health Organization*, League of Nations, Vol. VI, No. 5, 1937.

13. K. Stouman, "Health Indices for the city of Brussels," *Bulletin of the Health Organization*, League of Nations, Vol. VII, No. 1, 1938.

14. W. Selleslags, "Application des indices sanitaires des districts ruraux à la province de Luxembourg (Belgique)," Document C.H. 1946 de l'Organisation d'hygiène, Société des Nations, 1939.

15. World Health Organisation, *Chronicle of the World Health Organisation*, Vol. 1, Nos. 1, 2, 1947, United Nations.

Résumé

Malgré un long emploi de statistiques démographiques comme moyen de connaître a l'état sanitaire de la population, la comparaison des statistiques sanitaires continue de n'être pas satisfaisante. Les principales difficultés sont: l'absence des chiffres indispensables dans certaines régions; la diversité des méthodes utilisées pour leur compilation ou clasification; les divers principes d'interprétation; le nombre excessif d'institutions nationales s'occupant de statistiques démographiques; les différents intérêts et objectifs servant de motif aux activités de chacune de ces institutions; et, dans certains cas, le défaut de normes appropriées ou même l'absence de lois appropriées dans le domaine statistique.

La relation entre la science médicale et la statistique avec la sociologie étant de jour en jour plus étroite, les indices sanitaires deviennent fort utiles pour procurer une idée exacte du rapport entre l'état sanitaire de la population et le milieu géographique, économique et social.

L'utilité d'indices sanitaires est bien prouvée par les services très efficients rendus, aux Etats-Unis, par des indices établis par l'"American Public Health Association".

En 1938, une commission spéciale des indices sanitaires de l'Organisation d'Hygiène de la Société des Nations, a choisi une série de 52 indices en vue d'établir une norme internationale comprenant des indices de vitalité et de santé, des indices du milieu et des indices caractérisant les services de santé publique.

Le but de ces indices sanitaires était de former un aperçu complet de la santé publique, considérée sous les trois principaux aspects: santé et vitalité individuelle, facteurs environnants et systèmes de campagnes sanitaires. A ces fins, le questionnaire contient des questions concernant les domaines suivant: démographie, mortalité, morbidité, météorologie, habitation, hygiène, consommation, moyens de communication, instruction publique, personnel et matériel médicaux, activités médico-sociales et de prévoyance sociale.

Toutefois, la seconde guerre mondiale empêcha de continuer le développement des indices sanitaires en Europe occupée. C'est donc sur le continent américain que nous pouvons trouver de la continuité dans le dévéloppement des indices sanitaires. Il a été possible aux Etats-Unis d'améliorer les indices sanitaires et l'Amérique Latine a fourni plusieurs contributions depuis 1940. Cette année-là, la délégation brésilienne à la quatrième conférence Pan Américaine de Directeurs de la Santé publique a proposé l'introduction d'un nouveau questionnaire pour les indices sanitaires, comportant 72 questions.

Quels ont été les enseignements de l'expérience en Amérique Latine? Nous avons appris au moins une chose: l'utilité d'indices sanitaires comme complément des statistiques démographiques. L'expérience de l'Amérique latine a aussi fait connaître quelques difficultés que l'on rencontre dans la pratique.

Le besoin d'un aperçu général sur le problème apparait d'autant plus clairement qu'une connaissance plus large des fonctions sociales de la médecine rompt avec la distinction faite généralement entre la médecine curative et préventive ou la médecine individuelle et collective.

Les résultats obtenus par ces expériences nous permettent de comprendre les principaux problèmes sanitaires d'aujourd'hui. Le premier de ces problèmes semble être d'élaborer un index ou des indices pouvant être universellement appliqués, en tenant compte des différents standards de vie dans les divers pays.

Discussion

Mr. Emerso stated, in his discussion of Mr. de Viado's paper, that it seems to him necessary to distinguish between care of the sick as a medical and social service through hospitals, etc., and as a public health service intended to protect and promote the health of the people. He remarked that indices of public health services are more difficult to make, and proposed that they should first be formulated by medical administrative health officers and teachers of public health practice, and later tested practically in collaboration with statisticians.

RECENT DEVELOPMENTS IN APPLICATIONS OF STATISTICS TO TUBERCULOSIS CONTROL PROGRAMS

by Jacob Yerushalmy

Principal Statistician, Tuberculosis Control Division
Public Health Service (United States)

The newer demands on statistics in tuberculosis control programs result primarily from a number of important recent technical developments and their application to the control of the disease. The more important of these developments are:

1. The development of photofluorography.
2. Newer knowledge of the pathogenesis of tuberculosis and a number of diseases which simulate it.
3. More widespread use of BCG vaccination against tuberculosis.
4. Development of antibiotics, especially streptomycin, which may be useful in the treatment of the disease.
5. In addition, more refined statistical methods have been developed to assist in specific epidemiological investigations of tuberculosis and its control.

It may, therefore, be well to review some of the recent developments in the applications of statistics to tuberculosis control programs in the different fields of activity mentioned. Because of space limitation it is not possible to present detailed discussions on the specific statistical techniques utilized or developed in the different projects. This report must, therefore, be limited to a bare outline; more details may be found for many of the activities in the reference articles. In addition, certain activities are mentioned for which the detailed statistical techniques have not yet been completely developed. Many of these will, no doubt, be reported in the literature in the future, when completed. It was thought desirable, however, to indicate in this report various types of activity which present different statistical problems even if these problems have not, as yet, been solved.

PHOTOFLUOROGRAPHY

The development of photofluorography provided a technique for utilizing X-rays in large scale tuberculosis case-finding programs. Within the last decade a number of inexpensive methods have been developed for obtaining chest X-ray films. These are photofluorography on 35mm., 70mm., and 4" x 5" film sizes, and roentgenography on 14" x 17" paper negatives. All of these techniques employ automatic devices which permit large savings in personnel and material, so that the X-raying of large groups, if not the majority of the adult population, is a real possibility. Indeed, public and private agencies in many countries are taking advantage of this development to carry on this activity on a very large scale.

Statistical activities which grew out of these extensive activities are numerous. The more important of them are associated with:

1. Evaluation of the diagnostic efficiency of various photofluorographic methods.
2. Interpretation and classification of X-ray shadows.
3. Prognosis of persons with shadows discovered in mass X-ray surveys.
4. Problems in administrative statistics which concern the follow-up of persons with X-ray shadows discovered in mass surveys.

Evaluation: In evaluating the diagnostic efficiency of different photofluorographic methods, it is necessary to obtain two measures for each technique:

A measure of sensitivity or the probability of correct diagnosis on the positive[1] cases, and

A measure of specificity or the probability of correct diagnosis on the negative[1] cases.

Measures of sensitivity and specificity for the different techniques are obtained by comparative studies whereby a sample of the population is X-rayed at the same time by means of photofluorographic technique and also by the conventional 14" x 17" celluloid film technique.

Two statistical problems present themselves in these studies. The first results from the fact that in addition to differences inherent in the techniques there are large subjective errors of interpretation. For example, even on the 14" x 17" celluloid films a competent radiologist or chest specialist may miss as many as 25 percent of the films which were called "positive" by another, equally competent, interpreter. Moreover, errors of almost the same magnitude are found in two independent interpretations by the same expert on the same set of 14" x 17" films.

The second difficulty results from the fact that it is not known who in the examined population is positive and who is negative. If one of

---

By "positive" cases are meant persons who have X-ray evidence suggestive of tuberculosis, and "negative" cases refer to persons who have no such evidence.

the techniques, for example, the 14" x 17" celluloid is accepted as a standard, this difficulty does not arise. However, in view of the relatively large subjective errors which exist also in the interpretation of the 14" x 17" celluloid films, it would appear more desirable to attempt an absolute comparison of the different techniques without reference to any of them as a standard. When such a comparison is attempted it is necessary to know who in the population has X-ray evidence of tuberculosis, but this information can be obtained only by the use of one or more of the techniques tested. It is, in a sense, a vicious circle.

A method of investigation needs, therefore, to be devised which would accomplish two things; it must provide an objective method of defining "positive" cases in the examined group, and it must minimize as much as possible, the effects of subjective errors of interpretation, and at the same time accentuate the differences which are due to the techniques themselves.

A recent investigation [1][2] succeeded partially in accomplishing these objectives. Individuals were defined as "positive" on the basis of evidence provided—not by a single technique, but by all techniques. Similarly, the interpretations of a number of competent interpreters were used in the definition rather than that of a single reader. Four different X-ray techniques were utilized in this study and the films obtained were interpreted independently by five different experts. These interpretations yielded 20 readings for each individual in the study; that is, each person had 20 chances of being called positive. This method allowed the classification of the persons examined according to different degrees of "positiveness" from zero to twenty. It was thus possible to stratify the examined population and to compare the number of positive readings obtained by each of the different techniques in the several categories of "positive" cases. The result obtained by this method of investigation was that all the techniques studied have equal efficiency in detecting cases of tuberculosis.

There are a number of limitations to this method of study and great care must be exercised in testing these limitations. Nevertheless, it seems to be an efficient and useful method of comparison which may be applicable also to fields other than X-ray interpretation where similar problems are present. Indeed Professor Neyman has recently developed a mathematical formulation which, when it is completed, may provide, a general solution applicable to many fields.

Interpretation and classification of X-ray shadows: Mass radiography is ordinarily conducted in population groups of apparently healthy persons. Consequently, the majority of cases discovered have no symptoms, and the only evidence of possible disease is that provided by a shadow on the X-ray film. In addition, the types of shadows found on these X-rays may be somewhat different from those with which the radiologists and chest specialists are familiar in their work in the clinic or sanatorium. As experience in mass radiography increased, it became apparent that the accepted methods of interpretation and classification are not entirely applicable to these shadows.

---

[2] Numbers in square brackets refer to references listed at the end of this paper.

Studies of a statistical nature were undertaken, first, to evaluate the accepted methods of classification and their application to mass radiography; second, to attempt to create different methods of classification of X-ray shadows which may be more useful in mass radiography even if they differ from those used in clinical work. These problems have, by no means, been solved to date; however, considerable progress has been made toward their solution. It has, for example, been definitely established that even an expert's single reading of a set of films may miss as many as 25 percent of the "positive" cases.[3] One practical result of this finding was the recommendation that every set of mass X-ray films be interpreted independently by more than one reader. In addition, it was shown that the usual classification of X-ray shadows into minimal, moderately advanced, and far advanced tuberculosis as well as their classification according to probable activity of the lesion or as to its morphology is not suitable for lesions discovered in mass radiography. No two persons will, in any reasonable way, agree on these classifications and even a single individual is not likely to be at all consistent with himself in two independent interpretations of the same set of films.

Experimentation with new methods of classification has not, so far, yielded positive results. Solution of these problems will depend on successful utilization of knowledge from a number of different disciplines such as radiology, pathology, clinical tuberculosis, statistics, physiological optics, and psychology. These studies are continuing now in the United States, but the solution is not expected for some time.

Prognosis: The extensive activities in mass radiography brought to light a large number of persons with X-ray evidence of tuberculosis who are symptomless. Since very little experience is available on which to base policy in respect to treatment, it became necessary to establish long-range follow-up studies of such individuals. These involve not only studies of correlation between the appearance of the shadow with clinical and laboratory findings but also studies whose aim is to determine whether prognosis of the cases can be predicted from the appearance of the shadow on the initial film.

---

[3] *I.e.*, cases which another reading of the same films will indicate as having X-ray evidence suggestive of tuberculosis.

A number of important statistical problems emerge from these follow-up studies. In the first place they need to be integrated with the theoretical studies mentioned above on interpretation and classification of lesions. In the absence of suitable methods of interpretation and classification, the studies were set up with the view of obtaining statistical answers to the problem of classification. In other words, simple methods of describing lesions and components of lesions are being developed. It is hoped that the significance of various appearances of a shadow will be brought to light in the course of the years of follow-up, when the future course of the lesion will be correlated with the descriptive terms which are being developed. There are also statistical problems in the quantitative assessment of changes in the appearance of a lesion on films which have been taken at different times.

In addition a number of important problems in the quantitative assessment of clinical and laboratory techniques associated with the follow-up of individual cases are emerging. These, obviously, also require statistical treatment for their ultimate solution.

In the main, the statistical activities in this field may be characterized as those of keeping pace with the dynamics of changing concepts in radiology and tuberculosis control. This may be accomplished by systematic collection of data on a large number of different characteristics and by devising methods of analysis which will allow the significant factors to reveal themselves.

Administrative statistics: Mass radiography activities also created a number of administrative problems which required statistical treatment. These are of two types: those associated with the smooth operation of the projects, and those related to the administrative follow-up of the cases discovered.

Since photofluorographic machines have a high degree of efficiency in examining large numbers of people in a relatively short period of time it became important to devise methods of recording the statistical data on individuals examined quickly and efficiently. The time necessary to record the information could not be greater than that required for the X-raying of the individual and yet the information must be sufficient to yield statistical data on the population examined and the amount of disease found. Various experiments with different methods of handling this problem were made and the methods now in practice make use of punch cards. These have been devised to serve several purposes such as index cards for future individual reference, mechanical tabulation for statistics, as well as for notifying the persons examined of the findings by utilization of an additional stub. There are, of course, a number of detailed problems associated with this activity, many of which may be found in an exposition describing a record system which was found to work satisfactorily. [2]

In the administrative follow-up of positive cases the problem was to devise means which would assist the administrator to ascertain that cases discovered in mass survey work are followed through. Since tuberculosis is a chronic and prolonged disease, the statistical tools necessary for the continuous follow-up of these cases are not of the simplest. The files need to be of a visible kind and the method of recording must be dynamic and adjusted to continuous changes in the status of individual cases. The installation and maintenance of such a system is a major operation and presents many interesting problems in administration and statistics. A recent manual for case record systems [3] provides one method of handling this problem.

PATHOGENESIS

X-ray examination of the chest constitutes only one of the tools used in the control of tuberculosis. The more important additional techniques are bacteriological examinations for the isolation of tubercle bacilli, and the tuberculin test. This latter is used not only for differential diagnosis, but also as an index of infection in the general population, and as such is very useful in public health activities.

In treatment of an individual case, as in the Public Health aspects of tuberculosis control, the interplay between the evidence provided by the X-ray and that furnished by supplementary tests creates a number of very important and interesting problems, especially when the results of the different tests are contradictory. It, thus, becomes necessary to sharpen the tools of diagnosis and to determine more exactly the specificity of the different tests used in tuberculosis control. These problems become especially acute in investigations on early development of tuberculosis and its ultimate course. For example, one of the major problems in tuberculosis is to determine the best test for detecting the development of a lesion in its incipiency.

It is clear that many answers to these questions can be obtained by extensive epidemiological investigations in which statistics is the main tool. Specific statistical problems, in this field of activity, are associated with those on:

1. Specificity of tuberculin tests
2. Development of early tuberculosis
3. Diseases simulating tuberculosis

Specificity of tuberculin tests: The reaction to a tuberculin test is presumed to indicate that the person has been, in the past, infected with tuberculosis. It does not necessarily mean that the person has, or ever had, clinical tuberculosis. It is, however, necessary to exercise great caution in the application and interpretation of this test. In the first place, there apparently are great variations in the response to different preparations of this antigen. Secondly, and most important, it is necessary to consider the dose to which a reaction has been obtained, and thirdly, confusion exists on the simplest matter of reading the reactions as positive or negative.

On the question of dosage it has been shown [4] that when the dose is not specified, reaction rates become meaningless. When a group of children had been tested consecutively by increasing doses of tuberculin it was found that nearly all reacted when the dose was sufficiently high. In other words, the notion of reaction rates has been changed to that of "sensitivity levels". It became possible to construct curves of the "S" shaped-cumulative type which measure sensitivity levels of different population groups. If detailed and exact information about previous exposure were available, this method could be used to determine the critical level below which the test is specific as a measure of
ve which the test is non-specific.
a band around the critical level is
exposure and infection history of
wn, the results are not sufficiently
reover, the problem still exists on
parations. In a sense it is similar
ent types of X-ray techniques. In
order to obtain useful information about tuberculin, it is necessary to know who in the population tested has been infected with tuberculosis.
, use must be made of
would seem that the
ined mainly by a care-
degree of reaction to
ng the response to the
idence, and by the si-
of different prepara-
in progress.

uberculosis: Tuberculosis in its early stages is usually asymptomatic. By the time symptoms develop the disease has, in most cases, advanced to such an extent that the chance of recovery of the patient is greatly reduced. In addition, the damage to the community in the form of transmittal of the disease to household associates and other contacts has by that time, for the most part, been accomplished. Consequently, it becomes important to study the early manifestations of tuberculosis and particularly to attempt to find easy criteria by which the disease may be discovered in its early stages. These considerations lead to seriatim studies of apparently healthy individuals from the time when they were free from tuberculosis to the early development of this disease. Such studies are being conducted on special groups of the population, particularly among student nurses. These groups are selected primarily because they constitute a relatively easy group to follow and because infection in this group is relatively frequent.

The statistical difficulties involved in such seriatim studies are well known. In this study the difficulties are exaggerated by the presence of all of the uncertainties concerning the tests used and the differences in the interpretation. For example, the problems of specificity of the tuberculin test and the difficulties in X-ray interpretation and classification are present to a very high degree. On the other hand, the homogeneity of the group as well as the additional knowledge relative to exposure which may be obtained in this group may be expected to assist greatly in the solution of the extraneous problems mentioned.

The greatest contribution which studies of this nature are likely to make results from the existence of data on changes from negative to positive. Since the individuals in such studies are examined [illegible] information becomes available on such matters as conversion from tuberculin negative to tuberculin positive, as well as on changes in the X-ray appearance. Statistical analyses of these are expected to provide information to answer such questions as: What are the earliest measurable manifestations of the development of a lesion? Is it a tuberculin conversion or is it the X-ray, or do both of these change at the same time? What are the intervals of time from infection to tuberculin sensitivity to the appearance of a lesion on the X-ray? What is the course of tuberculosis from the early stages to spontaneous recovery or to advanced disease?

Important by-products which may be expected from these analyses relate to the specificity of the tuberculin test and the correlation between that test and bacteriological and X-ray evidence of tuberculosis.

Diseases simulating tuberculosis: Because tuberculosis is asymptomatic for a large proportion of the cases, X-ray of the chest becomes the most potent tool in the control and management of this disease. There is the danger, however, that some persons may be diagnosed and treated as tuberculous when in reality they are not. This is a consequence of the fact that apparently there are other diseases which simulate tuberculosis in the X-ray appearance of the shadow. The existence of such diseases has long been suspected from evidence provided by the fact that there was a relatively large number of cases with X-ray evidence of tuberculosis, or of a previous infection with tuberculosis (calcified lesions), that did not react to the tuberculin test. It was, however, only recently that convincing evidence was produced which implicated two such diseases (coccidioidomycosis and histoplasmosis), or immunologically related diseases, as causing calcifications and possibly also X-ray shadows very similar to those produced by tuberculosis.

The demonstration of this fact was accomplished by means of a carefully conducted epidemiological-statistical study which made use of the facility provided by the studies on student nurses [5]. These nurses were tested at the same time with tuberculin and a number of fungus antigens, and by means of an X-ray of the chest. It was established that nearly all nurses with X-ray evidence of calcification reacted either to tuberculin or to histoplasmin. This finding seemed to indicate that there must exist a disease closely related to histoplasmosis, which is very prevalent in certain parts of this country. Considerable work has been done to determine the geographic prevalence of this condition and other interesting facts about it [6, 7]. A number of additional epidemiological-statistical studies in this field are now in progress.

IMMUNIZATION

[redacted] s to develop an
ade. The more
l Guerin—or in
ortunately, the
method of vacci-
whereby a num-
ber o tubercle bacilli,
causing a large number of deaths.

Within the last decade, however, many countries have begun to use BCG vaccination on a relatively large sc d it appears that its application is relatively safe. On the
[redacted] side, however, no clear-
trolled studies have been
tradictory, and in many
ability of the study and
was accepted on its face value and ex-
have already been undertaken, it is pro-
o consider evaluation studies. In other countries, such as the United States, where vaccination programs have not yet been applied widely, it may be hoped that the situation will permit evaluation studies to proceed simultaneously with the development of a service program.

It may be said, without hesitation, that the success of such a combined research-service activity depends to a great extent on the application of statistical methodology in the development of the program.

It is obvious that in no country could a program of vaccination be retarded for five to ten years in order to allow controlled follow-up studies. In addition, it is important to determine just how long BCG vaccination provides protection, if at all. Some experts express the feeling that even if BCG provides protection in the early years, it might have the opposite effect in later years.

A plan has recently been proposed in the United States which not only combines the elements of service and research, but appears to have other attractive features in that the vaccination procedure may be evaluated without utilizing the expensive method of following individual cases.

The plan is to vaccinate a very large number of children in the country. However, it is not the intention to make this service available to everyone, but to vaccinate only children in selected ages. For example, it may be possible to vaccinate all the children born in even years, while those born in odd years will be left unvaccinated. If such a plan could be worked out on a large scale, it would not be necessary to follow individual children, but to obtain information on the immunity provided by BCG vaccination by a study of the tuberculosis mortality curve by single ages. If vaccination provides immunity, it would become apparent in a relatively short time that the mortality curve is not continuous, but dips and rises at alternate ages. The control will be perfect since all the children live in the same communities, subjected to the same environmental and other conditions. For each vaccinated age group, the experience in the previous year of the unvaccinated children who are one year older and the experience in the following year of children who are one year younger will serve as controls. It will also be possible to investigate whether the amount of protection provided is dependent on age of vaccination. In addition, barring very large scale population shisft, information will become available on the length of time protection is provided, since the mortality curve can be followed for many years to come.

Other statistical problems relating to the BCG vaccination program are associated with studies on specificity of the tuberculin test, since, presumably only non-reactors to tuberculin are to be immunized. Incidentally, the implication of this last statement is being subjected to investigation. There are, of course, other problems such as those on the standardization of the product used and on methods of application.

ANTIBIOTICS

Although the tubercle bacillus has been isolated, known and studied for over half a century, no specific therapy has been developed which is successful in curing tuberculosis in human beings. In the last few years, however, the antibiotic streptomycin has been developed and it is sufficiently promising as a cure for at least certain types of tuberculosis to justify extensive work on its evaluation and further development.

Evaluation studies on streptomycin are in many respects, similar to evaluation studies in other fields of medicine. There is, however, one complicating factor in tuberculosis which may not be present in many other diseases. This is inherent in the chronicity of the disease and the difficulty in evaluating its prognosis. If streptomycin were the "miracle drug," the benefits of which could be seen by the naked eye, this would cause little difficulty. Unfortunately, this apparently is not the case, at least in the present state of the development of streptomycin. A useful evaluation of this drug can be obtained only by means of well conceived and controlled studies. In such investigations, statistical methodology will have a great contribution to make. Fortunately, such a study is now being carefully planned in the United States.

Problems confronted in such a study are of two types: selection of samples for the study and control groups, and methods of evaluation. In the first type, methods must be devised whereby the study and control group will contain sufficient numbers of the different types of tuberculosis cases, and also that the control group be as similar as is humanly possible to the study group with respect to a relatively large number of characteristics.

In the evaluation part, all the problems mentioned previously as to the difficulty in X-ray interpretation and classification, particularly in evaluating changes in the X-ray appearance, as well as those associated with the evaluation of bacteriological examinations come to the fore. In addition very disturbing problems arise in connection with the toxicity of the drug and the possible permanence of these toxic effects. It may also be mentioned that there is evidence that tubercle bacilli may develop immunity to this drug. If this should prove to be true a number of important epidemiological speculations as to the possible effect of such a phenomenon on the control of tuberculosis are of far-reaching interest and importance.

It is obviously premature to say that these problems have been solved, but it is a wholesome sign that the problems are realized in the beginning of this investigation and not, as in many others, at the end of the study.

STATISTICAL TECHNIQUES IN EPIDEMIOLOGICAL INVESTIGATION

The modern concept of epidemiology is much broader than that implied by its name as the study of epidemics. Instead, it has come to mean the study of disease as a mass phenomenon leading to a better understanding of the natural history of the disease, its source, method of spread and control. As such, epidemiology of necessity utilizes statistical techniques and methods as a major tool in many investigations.

In epidemiological studies of acute disease, relatively simple statistical methods are, in the main, sufficient. However, in the epidemiology of chronic disease, such as tuberculosis, a number of complications arise which call for special statistical treatment. This may, perhaps, be illustrated by considering the concept of a *case fatality rate*, which measures the risk of death associated with the development of a specific disease. In an acute disease, such as pneumonia, the rate is straight forward and may be obtained with relative ease. All that is required is to observe a number of pneumonia cases for the period of time in which the disease is present and to determine the number of deaths occurring among them in the interval. In tuberculosis the concept of a case fatality rate becomes much more complicated and difficult even of definition. This is due to the duration of the disease which may, in some cases, be prolonged over many years, and to the wide range which may be found in the interval between onset of the disease and recovery or death. For these reasons, the concept of a case fatality rate must be changed to that of a *survival rate.*

Similar considerations apply to such concepts as *secondary attack rates.* These are very useful indices in epidemiological investigations, in that they measure the risk of acquiring the disease when exposed to a known case. The determination of these rates for acute diseases is generally accomplished by observing for a specific period of time all associates in households with a known case. Since, however, the length of time which household associates of a tuberculosis case need to be observed is relatively long, many complications arise. For example, if secondary attack rates are to be determined by age, and if the period of observation is longer than one year, the population exposed to the risk of acquiring the disease at a certain age cannot be determined as simply as in the case of acute diseases. Moreover, in actual practice the household group does not remain stable for a long period of time; many persons are removed from or added to the group.

To deal with problems of this type, Frost [8] has introduced life table techniques to solve many epidemiological problems in tuberculosis. This method allows for continuous adjustment required by the removal from or addition to the group under observation, as well as for the proper definition of the group exposed to risk according to different characteristics, by substituting person-time of exposure for the number of persons exposed.

Since Frost first used this technique, a number of other investigators [9, 10, 11] have applied the same technique for the determination of a number of indices which are important in epidemiological investigations of tuberculosis. Among these are survival rates of persons suffering from tuberculosis of different degrees of seriousness; secondary attack rates associated with sputum positive and sputum negative cases of tuberculosis; rates of break-down of persons with arrested or cured tuberculosis and survival rates of persons who have undergone rehabilitation after discharge from a sanatorium.

Another important contribution of Frost to the epidemiology of tuberculosis, also of a statistical nature, concerns the more comprehensive evaluation of age-specific tuberculosis mortality which is derived from a study of "generation" or "cohort" mortality [12]. Analysis of age-specific mortality curves at a given instant of time is not sufficiently comprehensive for an understanding of many aspects of infection and immunity, because the curve is a composite one and the previous tuberculosis experience of persons of a given age at one time is different from that of persons of the same age in another period of time. Frost, therefore, reproduced the previous tuberculosis history of a group of persons by determining the age-specific tuberculosis mortality curve of a generation born in a given year. For example, the tuberculosis experience of persons born in 1880 (cohorts of 1880) can be traced by assigning to them the age specific death rates from tuberculosis which were known to have operated at the corresponding periods of time. The type of curve which results is much different from the instantaneous curve and shows, for example, that although at present tuberculosis mortality in the older ages is higher than in the younger, the older persons of today have undergone, in previous years, higher rates and have had a peak in tuberculosis mortality, when they were 20 to 29 years of age. By comparing successive cohorts by chronological decades, it was possible for Frost to study physical changes in resistance and to indicate that age is a very important factor. The same method of study is being applied now to other problems in the epidemiology of tuberculosis, particularly those relating to questions of natural immunity.

Among the other contributions of statistics to tuberculosis control are its continuous uses in the day by day operations associated with the supervision of cases, and in the evaluation of the tuberculosis problem and its control. Mortality and morbidity statistics are the best measures available for the latter purpose. It should, however, be mentioned that both morbidity and mortality reporting are not as complete and accurate as would be desired and that there are a number of differences in the completeness and accuracy of reporting for the different countries. This makes comparability on an international scale difficult. It is, however, encouraging that although there have, no doubt, been improvements in completeness of reporting, the reported death rates from tuberculosis until the outbreak of the last war have been declining in nearly all countries of the world. It is also noteworthy that there is a continuous change in the form of the tuberculosis mortality curve by age from one which had its peak at ages 20-30, to one which is continuously rising with age.

The recent devastating war had a profound effect on the tuberculosis mortality rate in many countries which were directly affected by the war. Although figures are fragmentary, there seems to be little doubt that many of the gains made in the control of the disease have in some countries been wiped out and in a number of countries tuberculosis is now as rampant as it was more than half a century ago.

The current evaluation of the course of tuberculosis mortality in nearly all countries is complicated by the very extensive population dislocations which have resulted directly or indirectly from the war. Very few countries have sufficiently detailed statistics on the population to determine age-specific death rates from tuberculosis. Because of this, there has been a greater tendency in recent years to utilize death ratios, or the ratios of deaths from tuberculosis to deaths from all causes, as a rough measure for at least the direction of the course of the disease. Such ratios have, of course, a number of very important limitations and great care must be exercised in their application. They need to be constructed specific for age and sex, and the other diseases must be studied concurrently to determine whether changes in this ratio may be due to extensive changes in frequencies of deaths from other diseases.

No attempt has been made in this brief review to encompass all the applications of statistics to tuberculosis control programs. It is also not claimed that the subjects presented are the most important; they only happen to be those with which the author is intimately acquainted. No doubt, many important activities and contributions in the United States and in other countries deserve equal, if not greater, emphasis. When this subject is viewed as a whole, it may be stated taht the attack on tuberculosis is proceeding in a much more organized and planned fashion than ever before and that statistical techniques and methods are contributing in no small measure to the better understanding of the disease and to a more efficient utilization of available resources in the fight against it.

References [4]

[4] It is not claimed that this list is, in any sense, complete. It is presented only as references to some of the subjects discussed in the text.

(1) C. C. Birkelo, W. E. Chamberlain, P. S. Phelps, P. E. Schools, D. Zacks, and J. Yerushalmy, "A Comparison of the Effectiveness of Various Roentgenographic and Photofluorographic Methods," *Journal of the American Medical Association*, Vol. 133 (1947), pp. 359-365.

(2) H. E. Hilleboe and R. H. Morgan, *Mass Radiography of the Chest*, The Year Book Publishers, Inc. (Chicago, 1945), Chapter 10.

(3) *State Central Case Record Systems and Local Case Registers for Tuberculosis*, U. S. Public Health Service, 1947.

(4) M. L. Furcolow, B. Hewell, W. E. Nelson, and C. E. Palmer, *Quantitative Studies of the Tuberculin Reaction, I, Titration of Tuberculin Sensitivity and its Relation to Tuberculous Infection*, Public Health Reports, Vol. 56, No. 21 (May 23, 1941), pp. 1082-1100.

(5) C. E. Palmer, *Nontuberculous Pulmonary Calcification and Sensitivity to Histoplasmin*, Public Health Reports, Vol. 60, No. 19 (May 11, 1945), pp. 513-520.

(6) C. E. Palmer, *Geographic Differences in Sensitivity to Histoplasmin among Student Nurses*, Public Health Reports, Vol. 61, No. 14 (April 5, 1946), pp. 475-487.

(7) S. H. Ferebee and M. L. Furcolow, *Histoplasmin Sensitivity among Siblings*, Public Health Reports, Vol, 62, No. 23 (June 6, 1947), pp. 834-847.

(8) W. H. Frost, " Risk of Persons in Familial Contact with Pulmonary Tuberculosis," *American Journal of Public Health*, Vol. 23 (1933), pp. 426-432. Also K. F. Maxcy, Ed., *Papers of Wade Hampton Frost, M.D.—A Contribution to Epidemiological Method*, The Commonwealth Fund (New York, 1941), pp. 582-592.

(9) Ruth R. Puffer, H. C. Stewart, and R. S. Cass, " Tuberculosis Studies in Tennessee; Subsequent Course of Cases Observed in Williamson County," *American Journal of Hygiene*, Vol. 28 (1938), pp. 490-507.

(10) H. E. Hilleboe, *Post Sanatorium Tuberculosis Survival Rates in Minnesota*, Public Health Reports, Vol. 56. (1941), pp. 895-907.

(11) L. E. Stiltzbach, " The Sheltered Workshop in the Rehabilitation of the Tuberculous; Medical Experience, " *Milbank Memorial Fund Quarterly*, Vol. 21 (1943), pp. 80-101.

(12) W. H. Frost. "The Age Selection of Mortality from Tuberculosis in Successive Decades," *American Journal of Hygiene*, Vol. 30 (1939), pp. 91-96, Also Maxcy, *op. cit.*, pp. 593-600.

Résumé

Dans la présente communication sont discutées les plus nouvelles exigences posées à la statistique par les techniques d'origine récente dans la lutte antituberculeuse telles que la radio-photographie la vaccination antituberculeuse au moyen du B.C.G. l'emploi de la streptomycine, et les nouvelles connaissances sur la pathogénèse de la maladie,

Parmi les problèmes statistiques que l'on peut aborder dans ces activitiés beaucouptressemblent à ceux des autres domaines d'application d'autres sont plus spécifiques. A titre d'exemple de ceux-ci on peut citer celui du diagnostic médical. que lon trouve dans les études comparatives sur l'évaluation de l'efficacité des diverses techniques de la radiophotographie et de la radiographie.

Deux problèmes d'ordre statistique se présentent dans ces études. L'un résulte du fait qu'en dehors des différences inhérentes aux techniques mêmes, il y a d'importantes erreurs subjectives d'interprétation. Par exemple; même sur les clichés à 35/42.5 cm. il est probable qu'un radiologue ou un clinicien phtisiologue avisé laissera passer inaperçu jusqu'à 25 p. 100-des clichés qu'un autre, également compétent. a désignés comme "positifs". La seconde difficulté résulte du fait qu'il n'est pax connu quels membres de la population examinée sont positifs et lesquels sont négatifs. Etant données les erreurs subjectives assez inportantes, qui existent aussi dans l'interprétation des clichés à 35/42.5 cm., il paraît désirable d'essayer une comparaison absolue des diverses techniques sans choisir aucune d'entre elles comme norme. Si l'on aborde une telle comparaison il sera nécessaire de savoir quelles personnes dans la population, présentent des lésions tuberculeuses décelables par les examens radiologiques. mais on ne peut se renseigner sur ce point que par l'emploi d'une ou de plusieurs des techniques qu'on voudrait mettre à l'épreuve. C'est un cercle vicieux, en quelque sorte.

Il sera donc nécessaire de trouver une méthode d'investigation à deux fins: 1) fournir une méthode objective pour définir les cas positifs du groupe examiné, et 2) réduire. autant que possible, les effets d'erreurs subjectives d'interprétation, et d'autre part. accentuer les différences attribuables aux techniques mêmes.

Une investigation récente a réalisé ces buts en partie. On a classé les cas comme "positifs" selon l'évidence fournie par l'ensemble de toutes les techniques, non pas par une seule technique. D'une manière analogue, au lieu de l'interprétation des résultats par un seul expert, les interprétations de plusieurs experts ont été utilisées pour le classement. Dans l'étude en question, on a employé quatre techniques différentes de radiologie, et cinq experts ont évalué indépendamment les clichés ainsi obtenus. Cela a fourni vingt interprétations pour chaque individu compris dans l'étude; c'est-à-dire que chaque personne avait vingt chances d'être désignée comme "positif". On peut classer les individus examinés d'après les différents degrés de "positivité" de zéro jusqu'à vingt. On a donc pu stratifier la population examinée et comparer le nombre d'interprétations positives obtenues au moyen de chacune des diverses techniques dans les catégories respectives de cas "positifs".

Cette confrontation a montré que toutes les techniques étudiées possèdent une efficacité égale dans le repérage des porteurs de lésions tuberculeuses.

Cette méthode d'investigation a un certain nombre de limitations, et il faut apporter un soin extrême à l'analyse de ces limitations. Néanmoins, elle paraît une méthode aussi utile qu'efficace, une méthode qu'on pourra peut-être appliquer à des domaines autres que l'interprétation radiologique, dans lesquels des problèmes semblables se présentent. En effet, la forme mathématique que *M. Neyman* en a donnée fournira peut-être, quand on l'aura completée, une solution générale applicable à beaucoup de domaines.

Discussion

Mr. Neyman's contribution to the discussion, a short paper on the statistical treatment of the problem of diagnosis, was read in French by Mr. Darmois. This short outline follows, with an example of its application omitted :

"Every attempt at a mathematical treatment of phenomena must begin by building a simplified mathematical model of the phenomena. In studying the problem of diagnosis as presented by Mr. Yerushalmy, we have to consider a population (or a universe) U of individuals which we shall imagine divided into three exclusive categories: (i) Individuals *entirely free* from the given disease, whose proportion is α, (ii) Individuals *moderately affected* by the disease, whose proportion is β; (iii) Very heavily affected persons, whose proportion is $\gamma = 1-\alpha-\beta$.

"Mr. Yerushalmy's data and the discussion suggest that a diagnosis is always subject to error. Therefore we have to consider the probabilities of a correct outcome of the diagnosis. The probabilities are different for each of the three categories of individuals distinguished above. Let then π and ρ stand for the probabilities of correct diagnosis for categories (i) and (ii) respectively. In consequence $1-\pi$ will denote the probability of a healthy person being classified as ill or, to use Mr. Yerushalmy's expression, the probability of a "false positive." The probability of a correct diagnosis for a very ill person is postulated to be unity.

"The mathematical model outlined above contains four independent parameters with which to describe the machinery behind figures such as those collected by Mr. Yerushalmy: the proportions α and β of the the population and the probabilities π and ρ of a correct outcome of diagnosis.

"In order to be able to estimate all the parameters in a mathematical model it is necessary to write down an equal number of independent equations and this implies at least an equal number of observed independent frequencies. Moreover, should it not only be necessary to estimate the unknown parameters, but also to test some hypotheses concerning them, the number of independent frequencies will have to be larger than the number of parameters. Thus if we assume the above model with four parameters and desire to test some hypotheses concerning it, then the necessary data would have to involve at least five independent frequencies.

"Let n denote the number of *independent* diagnoses made for each individual of the N individuals constituting a sample from the universe U. Denote by k the number of positive diagnoses made for a particular individual. Obviously k may have one of the $n+1$ values 0, 1, 2, ..., n. Let N_k stand for the number of individuals in the sample for which there were exactly k positive dagnosis. Dividing $N_k/N = q_k$ we obtain the relative of exactly k positive diagnoses in the sample. There are $n+1$ such frequencies but only n of them are independent since they must add up to unity.

"It follows that, if it is desired to test some hypotheses concerning the four parameter models described above, the number of independent diagnoses made by the same method for each individual of the sample must be at least equal to five. If this condition is satisfied, then the procedure of estimating the parameters α, β, π, and ρ may be that of the best asymptotically normal estimates, explained in a recent paper.[1]

"The possibility of applying this method to data collected by Mr. Yerushalmy depends on whether or not one agrees to consider that these data are diagnoses, *by the same method, independently applied to the same individuals....*"

Mr. Yerushalmy pointed out, with reference to Mr. Neyman's paper, that the method was not quite complete. In the comparison of several techniques, even though it can be shown that, e.g., each technique diagnoses the same percentage of positive cases, essentially the method does not prove that the techniques are equally good, since there must also be success in diagnosing the right cases.

---

[1] J. Neyman, "Contributions to the Theory of χ^2 Test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, 1945, 1946.*